INTEGRATED TRAFFIC CONTROL MECHANISMS FOR ATM NETWORKS

By

SHANG-YI LU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1994

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the members of my supervisory committee for their kind help and guidance throughout this work. I would like to express special thanks to my committee chairman, Dr. Haniph A. Latchman, for his invaluable suggestions and sincere instruction through this research. I would also like to thank especially Dr. Scott Miller, Dr. Randy Chow, Dr. Leon W. Couch II, Dr. Donald G. Childers and Dr. Yang-Hang Lee for discussions, corrections, and encouragement during all these years. Special thanks should be given to Mr. Bill Waggener and Mr. Nimish Shah from Loral Data Systems for their kind support for the whole project and invaluable suggestions and discussions throughout the research period.

I would like to express my appreciation to Ms. Pam Mydock for her excellent work in proofreading this dissertation. I would like to give special thanks to my colleagues and friends who encouraged me to go on for my graduate career: Wenyen Fu, Attique Ahmad, Diane Warfield, Kwang Rip Hyun, Azhar Khan, John Miller, Ron Smith and Abdul Majid Khan. Finally I would like to thank my husband for accompanying me through the years without any complaint and my parents, brothers and sisters for their full support, both spiritually and financially.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INTEGRATED TRAFFIC CONTROL MECHANISMS
FOR
ATM NETWORKS

By

Shang-Yi Lu

December 1994

Chairman: Dr. Haniph A. Latchman
Major Department: Electrical Engineering

Asynchronous Transfer Mode (ATM), a new cell-based transport technology, provides a flexible means to multiplex and switch variable rate information with diverse traffic characteristics and service demands on a single unified network. However, the need to accommodate the large spectrum of potential applications creates new challenges in the design of ATM switches and effective traffic control mechanisms. The most critical issue for a successful deployment of the ATM technology is to design packet switches capable of switching relatively small packets at extremely high rates while supporting traffic management functions to provide adequate service guarantees.

In this research an integrated traffic control framework for various ATM service categories and performance objectives is proposed. A switch model which accurately reflects the hardware design of an ATM fast packet switch has been developed as an

experimental platform for evaluating the feasibility and performance improvement with different traffic control mechanisms. Several effective traffic control schemes have been presented and optimized with respect to critical network resources within a complete network environment as well as on a single switching node. We propose a generalized Partial Buffer Sharing (PBS) scheme to manage effectively the finite buffer at a switching node. A queueing model has been developed to characterize analytically the multiple-threshold system and the optimization procedures that dimension the system efficiently to achieve an optimal performance have been presented and verified. In addition, we conduct a thorough investigation on two effective flow control mechanisms, the Backward Explicit Congestion Notification (BECN) and the credit-based schemes, for ATM best-effort traffic. While the credit-based scheme gives a superior performance, its complexity for practical implementation is often prohibitive. On the other hand, our results show that a modified slow-start BECN scheme can be designed to attain asymptotically the same level of performance as the credit-based scheme, with a significant reduction in complexity. This research work contributes to the ATM networking technology with higher resource efficiency, improved network performance, and better resource protection.

CHAPTER 1
INTRODUCTION

1.1 Background

The development of high-speed integrated communication networks has reached a point where switching systems rather than transmission systems are the bottleneck for growing traffic rate and widely ranged applications. In an effort to develop new switching technologies for Broadband Integrated Services Digital Network (BISDN), the Asynchronous Transfer Mode (ATM) has been chosen by International Telegraph and Telephone Consultative Committee (CCITT) as the transport mechanism for a variety of classes of digital data. ATM is a statistical multiplexing and switching technique which is based on fast packet switching concepts. The information flow within ATM networks is organized into fixed-size packets called cells. Thus the ATM technology allows different traffic types to be multiplexed in a flexible manner and offers a potential efficiency improvement over synchronous transfer mode through the statistical sharing of network resources by multiple ATM connections.

BISDN will support a wide variety of network traffic and services with diverse characteristics and performance objectives. In particular voice, video and data traffic will be carried by the broadband network in an integrated fashion. Examples of potential application areas include high-speed interconnection of existing data networks, distributed file and procedure access, video telephony, computer imaging, and multimedia broadband services such as interactive high-resolution image communications. Indeed, it has been noted that besides the increasing demand for high-speed data communications, the emerging market for digital image services has become a significant factor driving the evolution of high-speed transport services [1, 2, 3].

Another area of major importance is the deployment of ATM technology and standards in campus, backbone networks and local area networking (LAN) environments, the so-called ATM LANs [4, 5]. It is envisaged that the new ATM LAN technology will overcome the deficiencies of older LAN technologies such as Ethernet and Fiber Distributed Data Interface (FDDI), which generally cannot provide the necessary bandwidth and performance guarantees to applications and individual hosts and also cannot support the increasing user population. The important advantages that ATM LANs offer over competing technologies include high-capacity networking, the ability to handle multiple traffic types, the flexibility to cater for different link speeds, and standardized access to broadband public network services when BISDN becomes available.

The design of an ATM switch is of fundamental importance to provide the connectivity which will support all these diverse applications. These potential services require bandwidths ranging from a few kilobits per second to several hundred megabits per second and qualities of service ranging from strict network performance guarantees to *best-effort* service class with little or no performance guarantee. Moreover, many of the traffic sources are highly bursty, and may generate cells at their peak rates for short periods and then become inactive. The heterogeneity in the requirements and traffic characteristics of these different services have important influences on the ATM switching technology. The key objective is to design and build packet switches capable of switching relatively small packets at extremely high rates while supporting traffic management functions to provide the necessary performance guarantees.

1.1.1 Fast Packet Switching

Packet switches were originally restricted to pure data communication applications, due to the speed limitation and the complexity of the required control protocols. However, with the introduction of high-speed and low-error-rate digital transmission

systems, the communication protocols used in conventional packet switches were substantially simplified. These simplifications made feasible the construction of high speed hardware-based processing and switching modules, thus increasing the range of potential applications of packet switches. Fast packet switching has emerged as a promising switching technique to support the wide range of communication services envisaged for BISDN. Considerable effort has been devoted to the development and analysis of fast packet switches in recent years. In fast packet switching technology, the transmission facility is used as a "digital pipe" to carry short packets of information one after another. Information in the header of each packet identifies to which of many logical connections the packet belongs. Fast packet switching utilizes the statistical multiplexing technique to provide flexibility for handling variable rate connections and achieving high utilization of network resources. With this multiplexing scheme, connections of arbitrary bandwidth are accommodated in a simple and natural way.

1.1.2 Need for Advanced Traffic Control Mechanisms

With communication networks operating at a much higher speed than traditional data networks, the design of an effective traffic management and congestion control scheme becomes one of the fundamental issues. Most conventional packet switched networks carry only nonreal-time data traffic which can be flow-controlled reactively in case of network congestion; such traffic control schemes are generally acknowledgment-based. They react to network congestion by instructing the traffic sources to throttle their excessive traffic flow (e.g., reduce their flow window). However, due to the increased ratio of propagation delay to cell transmission time, these reactive controls which depend on simple feedback signals from the receiver may be very inefficient for ATM networks. A large amount of information can be in transit within the pipeline by the time that feedback reaches the traffic source.

Applying purely reactive controls can result in a vast loss of information during network congestion and induce data retransmissions which will further deteriorate the situation.

Another challenge arises from the need to handle a wide variety of network traffic with diverse characteristics and performance objectives via a common high speed switching architecture. In particular ATM switches are expected to support very bursty and highly unpredictable data sources along with predictable and smooth Constant Bit Rate (CBR) sources and with other relatively smooth traffic. In the past those different applications were typically transported and switched on separate networks and thus networks could be optimized individually to meet the different performance objective for each application.

The integration of a wide diversity of traffic and Quality Of Service (QOS) constraints in ATM networks calls for new switching architectures and control schemes. Real-time traffic, such as voice and video, is mainly delay-sensitive and cannot be recovered through retransmission; these traffic sources have stringent performance requirements in terms of delay, delay jitter and cell loss. On the other hand, conventional data sources, depending on their applications, have diverse network performance requirements. For instance, file transfers are not usually delay sensitive but are loss sensitive, while remote login and telnet data applications are sensitive to absolute delay. The network will not be able to guarantee the promised QOS unless adequate control is exercised on critical network resources (i.e., switching and communication bandwidth, buffer space and processing capacity). The role of resource management and congestion control is obviously more involved in an ATM environment than in a conventional data network. In particular, the control schemes implemented in an ATM network need to be effective in yielding predictable and reliable network performance and must also be flexible in accommodating different traffic and service demands.

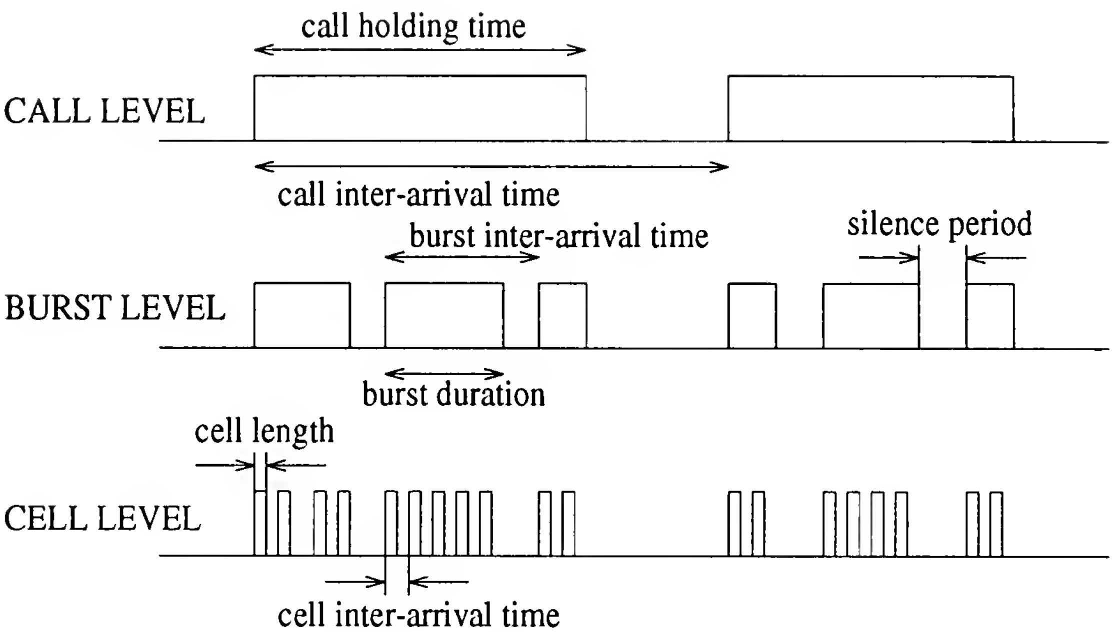


Figure 1.1: Multilevel traffic processes

Recently, traffic and congestion control in ATM networks has received a tremendous amount of attention. A variety of traffic control strategies have been proposed over the past years. Among them, multilevel control frameworks have been proposed by several researchers and considered to be an effective and efficient approach to prevent network congestion and to handle different service demands [6, 7, 8]. The multilevel traffic and congestion control is motivated by the fact that traffic sources usually have traffic states characterized by different levels, namely, call level, burst level and cell level. As shown in Figure 1.1, calls are composed of bursts which are, in turn, composed of cells. Therefore, congestion phenomenon should be evaluated and controlled at different levels and time scales. Hierarchical layering of controls were proposed to prevent congestion through proper call management at network access nodes and through resource allocation within network elements. Additionally, traffic shaping mechanisms are performed at the network access nodes to smooth out traffic burstiness. The distributed source control scheme proposed in [6] and the congestion control framework described in [7] are examples of multilevel traffic control strategies.

Generally speaking, traffic control mechanisms used in virtual circuit based packet networks (such as ATMs) can be divided into three major components [9]: (1) call/session admission controls to ensure a small probability of congestion; (2) sophisticated network resource management schemes to provide different QOS as well as efficient use of network resources; and (3) reactive controls which are able to be invoked in real time to minimize adverse impacts of congestion and to guarantee fairness between competing users. The first two belong to avoidance controls and are aimed at preventing networks from reaching an unacceptable level of congestion by conservative admission policies and preventive resource controls. Although these control strategies cause more processing overhead and thus reduce utilization of network bandwidth, they provide a robust means to support diverse applications and to guarantee QOS for critical traffic in broadband integrated networks. Incorporating congestion avoidance controls along with reactive control mechanisms as backups is considered to be the optimal approach to solve the network congestion problem.

1.2 Problem Statement

The multilevel traffic and congestion controls provide a framework that allows effective management of congestion occurring in different time scales. Many research efforts have been directed toward the development of various traffic admission control algorithms, applied at network access nodes to manage traffic loading. However, the issue of network resource management and maintenance, which is equally essential for the multilevel traffic controls to be successful, has not yet been well addressed.

The capability of a network element to engineer critical resources adequately in both normal and overload conditions is necessary and crucial for ATM networks. Our interest in this research is motivated by the following reasons: (1) Many data sources are very bursty and highly unpredictable and they are very difficult to be accurately characterized in advance of transmission. Strict access controls result in

low efficiency and are not well-suited for this type of traffic. More dynamic controls functioning at network elements as well as traffic access nodes are needed to achieve high resource efficiency and ensure adequate protection. In addition, some of traffic sources are not flow-controllable, for instance, voice and video traffic sources cannot stop generating cells even when the network is congested. Real time controls at network elements are important to prevent severe performance degradation during network overload; (2) Due to the interaction between different traffic streams within a packet network, packet flow can cluster together and create local congestion, even when the arriving traffic streams are well-regulated at their access nodes. Study has shown that the worst case loss and delay performance of a packet network may not be sufficiently controllable from the edges of the network [10]; and (3) As has frequently been observed in operational networks, network users may sometimes misbehave. A user may transmit data at a high rate without listening to the network control information. Moreover, such misbehavior can also be caused by software or hardware failures or by protocol implementation errors. Some access control schemes, such as virtual leaky bucket, would let the excessive cells get through at risk, that is, by tagging excessive traffic cells so that they can be discarded in case of congestion. It is the responsibility of the network control to prevent misbehaving users from interrupting normal service to others.

In addition, over the past years, the research works in the areas of ATM switch design and traffic management have progressed essentially independently [11]. Few of the previous studies of ATM traffic control really take into account the feasibility and complexity of implementing their control strategies on a realistic switch structure. In a realistic switch, the limited processing speed, finite buffer capacity and system architecture may become a bottleneck for implementing a complicated traffic control mechanism. The issue of the operational details of ATM traffic control schemes needs to be addressed carefully since it will not only directly affect the complexity of

network system hardware but also determine if ATM applications can be supported economically.

This research is intended to investigate effective integrated control mechanisms, as well as proper architectures, for switching nodes within an ATM environment. The integrated controls, when incorporated with a well-suited access control scheme, can be applied to ATM switches to enhance overall network performance and to deal with different service demands. The main objective of the proposed control mechanism is to provide adequate network resource management to support various QOS requirements and allow high resource efficiency.

1.3 Dissertation Outline

This dissertation is organized as follows. An introduction of the development and current status of ATM technology is first presented and addressed in this chapter. The critical issues for the successful deployment of ATM technology are also discussed.

In Chapter 2, an overview of previous literature in this area of research is given. This overview covers many research works that have been conducted in earlier stage and their pros and cons are discussed.

In Chapter 3, we motivate our considerations in this dissertation of a shared medium/output buffering ATM switch. This study is part of the research project supported by Loral Data Systems. It contributes to the development of the Loral fast packet switch and, in addition, provides a close insight into the ATM switch design. This research has considered three major approaches to ATM switch design: shared medium, shared memory and space division. A complete analysis was performed of the ATM switching technology. The performance characteristics, the limitations and the potential application areas of different switch architectures were investigated thoroughly. We developed a simulator which reflects accurately the hardware design of the Loral CPS-100 ATM-based switch and conducted an exhaustive simulation study

to analyze the switch characteristics. We show that critical network performance can be controlled effectively with the implementation of appropriate traffic control mechanisms. The study provides a realistic ATM switch model, which is fully acceptable for supporting all types of ATM services at the preliminary stage of ATM networks, as an experimental test-bed to evaluate the feasibility and performance improvement with different traffic control strategies.

An integrated traffic control framework for ATM networks is presented in Chapter 4. A study is conducted of the traffic characteristics, QOS requirements, modeling methods and applicable traffic management schemes of potential ATM applications. Specifically, ATM networks are expected to support three basic types of services: Constant Bit Rate (CBR), Variable Bit Rate (VBR) and Available Bit Rate (ABR). To accommodate the wide variety of ATM applications with diverse traffic characteristics and performance objectives, dynamic controls need to be enforced on critical network resources to achieve high resource utilization and ensure adequate protection. Appropriate traffic management schemes should be applied to different service types based on their traffic characteristics and QOS and, moreover, they should be effective on different time scale levels to provide a robust and complete control. An ATM traffic control architecture that integrates various traffic management schemes with effective control functions is proposed.

In Chapter 5, the issue of managing optimally the finite output buffers of an ATM switch is investigated. This research considers a generalized space priority queueing strategy, the Partial Buffer Sharing (PBS) scheme, which provides preferential treatment to the traffic with sensitive loss requirements. A queueing model that characterizes analytically a multiple-priority shared buffer system controlled by the PBS scheme is presented. The queueing system is modeled as a discrete time Markov chain with finite buffer capacity B and N classes of arrivals. With the assumption of a multinomial traffic distribution, a queueing model that predicts accurately the

steady-state loss probabilities of a multiple-class system under a set of given loss thresholds is developed. In addition, numerical procedures are introduced to optimize the threshold level selection to yield the maximal system admissible load. By employing the optimization procedures, the design space can be explored efficiently to find the best thresholds without making an exhaustive search over the entire range of possibilities. Numerical examples are provided to show the verification and effectiveness of the optimization procedures. The objective of the study is to dimension the finite buffer space properly to guarantee acceptable loss probabilities for different service classes. Making use of the priority control, the system is able to support different grades of loss quality with a minimum hardware cost and also provide a better control on the delay introduced during the buffering. It is shown that system performance can be improved substantially by dimensioning the buffer capacity to a moderate size and imposing appropriate threshold levels.

While Chapter 5 investigates the priority control scheme that manages critical resource on a single switch node, in Chapter 6 the issue of controlling traffic flows within a communication network is discussed. Although priority control provides an effective method to manipulate finite network resource as congestion occurs, priority control doesn't have the ability to release networks from congestion status since it cannot reduce incoming traffic. This research considers the feedback flow control schemes that apply to the ATM best-effort service class to minimize adverse impacts of congestion as well as guarantee fairness between competing users. The study is motivated by the fact that the traffic characteristic of this type of service is very difficult to predict in advance of transmission and, therefore, strict admission controls are not well-suited for this traffic type. What is needed for these applications is a dynamic control that can be evoked in real time to regulate the traffic flow along a virtual circuit as necessary. Two classes of flow control which have been proposed to the ATM Forum are investigated: the credit-based and the rate-based flow control schemes.

This research gives an exhaustive comparative study of these two control schemes and addresses in detail their relative advantages and disadvantages. It is shown that with carefully designed system parameters, the control schemes can prevent network throughput from being severely affected by packet retransmissions and provide substantial performance improvement. We conduct a complete analysis on the issues of connection transient behaviors and fairness of resource sharing by performing the simulation on different network configurations. The credit-based scheme provides a superior performance but requires a significant amount of specialized hardware at the ATM layer. While with the rate-based scheme the users may suffer performance degradation under extreme variations of traffic load, the rate-based approach only imposes very moderate requirements on the switch hardware. We demonstrate that the performance of the rate-based scheme can be improved to achieve asymptotically the same level of performance as the credit-based scheme by making use of a slow-start procedure and a larger buffer.

Finally, in Chapter 7 a conclusion of the research work is drawn and some future research is proposed.

CHAPTER 2
PREVIOUS WORKS

2.1 Bandwidth Management in Broadband Network

There are a lot of research works being undertaken in developing the best approach to manage network resource and deal with congestion problem. Some important works that proposed to control congestion in a preventive way include admission control and bandwidth enforcement (e.g., equivalent bandwidth, leaky-bucket, distributed source control, framing strategy and fast buffer reservation), link-by-link flow control, priority control and traffic shaping. Some other works proposing reactive congestion control schemes include forward and backward congestion indication, adaptive window sizing, throttle controls and selective cell discarding. In this section, a review of the important previous research and a comparison of their features are presented.

2.1.1 Equivalent Bandwidth

When a call request is received, the network needs to determine whether to accept or reject the new connection, based on the traffic characteristics and the current network load. Each accepted connection is then policed during the data transfer phase to ensure that the admitted traffic conforms with that specified at call establishment. To make the decision, the network has to compute the amount of effective bandwidth that needs to be reserved for the new connection and evaluate the impact on the current network condition. Because of the potentially dramatic differences in the statistical behavior of connections, the problem of characterizing the equivalent bandwidth of a new connection creates a critical challenge and has drawn

lots of research attention. Various approaches have been proposed to determine the equivalent bandwidth of connections: some use sets of precomputed curves obtained by analysis or simulation [12, 13, 14, 15] while some address computational procedures based on suitable traffic descriptors and the desired QOS [16, 17]. The important considerations in choosing such an approach are if the traffic model can accurately reflect the dynamic nature of network traffic conditions and connection characteristics and if the computation is simple enough to be consistent with real-time requirements.

2.1.2 Leaky-Bucket Control Scheme

The leaky-bucket scheme proposed originally by Turner [19] is a bandwidth enforcement algorithm for controlling the average cell rate and the burstiness of a logical connection. Various versions of the algorithm have been proposed. The fundamental principle of the leaky-bucket scheme is that an arriving cell can be transmitted only if the corresponding connection has a token in its bucket. A token is consumed from the bucket when the user transmits a cell. If there are no tokens available, the cell can be discarded or be stamped for preferential discarding in the event it encounters congestion. Tokens for each connection are generated at an expected average cell rate and a certain number of tokens (up to the bucket size) can be saved so that the maximum burst size is controlled.

The leaky-bucket approach enforces traffic control at the network access node and no further control over the order of service is used once cells are accepted. Therefore priority control needs to be enforced on network elements to provide satisfactory network performance for different classes of traffic. It has been shown that the maximum delay in a packet network can be guaranteed by making use of fair queueing service discipline at traffic multiplexing points, in conjunction with the leaky-bucket admission policy [20]. The leaky-bucket control scheme allows performance guarantees to be traded off against resource efficiency and traffic burstiness. However, one of the drawbacks of this approach is that there are currently no computationally

effective ways to decide when a new connection can be safely multiplexed with other existing connections [21]. In addition, the network resource can be under-utilized if the policing policy specified by user is too conservative.

2.1.3 Distributed Source Control Scheme

Ramamurthy and Dighe presented a Distributed Source Control (DSC) algorithm for controlling the rate of traffic entry into a broadband network based on pre-negotiated control parameters [22]. DSC is basically a congestion avoidance control strategy that regulates traffic sources to conform to traffic metrics that can be supported by the network. The underlying philosophy behind DSC is that the function of congestion control in the network should be done with respect to the network time constant (NTC). NTC defines the interval over which network traffic must be averaged and is determined by the link bandwidth, the amount of memory at each node, and the delay requirements of different traffic types that are being integrated in the network. DSC involves the negotiation of two control parameters, W_s (the window size) and T_s (the smoothing interval), between the source and the network access node. The ratio, W_s/T_s is chosen to be equal to the average throughput expected by the source when active. It has been shown that the network delay and cell loss performance are improved by reducing the smoothing interval T_s. A multilevel congestion control strategy was proposed later on [6] based on this approach. Depending on the time constants of events that are being controlled, a different level of control is evoked to prevent network congestion. The approach involves complicated traffic control and parameter negotiation procedures, which may cause significant control overhead on the network. Moreover, since the traffic control scheme is performed only at the network access node, there is no evidence showing that the worst-case delay and cell loss probability are bounded.

2.1.4 Framing Strategy

Golestani proposed a time-framing congestion control strategy, based on a packet admission policy at the edges of the network, and a service discipline called Stop-and-Go queueing at the switching nodes [10, 24]. The underlying idea of the framing strategy is to confine packets within certain logical containers traveling in the network, called time frames, and, therefore, the original smoothness of the admitted traffic can be preserved throughout the network. The framing strategy ensures that the burstiness of the traffic anywhere inside the network corresponds only to the burstiness of the admitted traffic at the edge of the network and that it is not altered as a result of complex interactions between traffic streams inside the network. Thus bounded delay and delay-jitter of each packet can be guaranteed. This approach provides an attractive feature for traffic which is delay or delay-jitter sensitive, such as video traffic.

However, the benefits of congestion-free and bounded-delay transmission are basically accomplished at the cost of a strict admission policy to enforce the smoothness property on packet arrivals. In order to achieve a reasonable bound on the end-to-end delay and delay-jitter of each packet, we usually have to choose a small frame size which is typically only a small fraction of a second. Since such a small averaging period is often insufficient to smooth out the statistical fluctuations of traffic sources, this admission policy requires in practice that capacity be allocated based on the peak rate of the connections. Consequently, this approach potentially leads to low utilization of the network resource.

2.1.5 Virtual-Clock Traffic Control Scheme

The virtual-clock traffic control algorithm proposed by Zhang [23] is a rate-based approach to control congestion. In the virtual-clock strategy, a virtual-clock, which ticks at every cell arrival from a data flow, is assigned to each logical connection.

Initially, the virtual-clock is set to the real clock time. The tick step is set to the expected average inter-cell arrival time. Thus the value of the virtual-clock will denote the expected arrival time of the arrived cell. The multiplexing node stamps the cell from each connection with its virtual-clock time and orders cell transmission based on the stamp values. If a user transmits according to the specified average rate, the virtual-clock control assures that the user's request throughput is guaranteed, that is, the minimum throughput will not be affected by misbehaving users. The users who violate their reservation will receive the worst service since their virtual-clock advances so far that their cells will be placed at the end of the service queues. Note that the excess traffic would still get service if the resource is free. Consequently, the resource utilization is maximized while the interference among different users is prevented.

The virtual-clock strategy may be viewed as a generalization and abstraction of the round-robin service discipline. Its major advantage is the provision of fairness and minimum throughput guarantees in the services offered to competing connections in the network. A drawback of this approach is that no bound has been obtained currently for the delay and jitter performance associated with it. This approach may not provide a satisfactory performance guarantee for the traffic which has stringent requirements on the worst-case delay and delay-jitter. Another problem that has to be considered for the virtual-clock control is the complexity and cost of implementation on a traffic multiplexing node. Since a cell that arrives late may be stamped with an earlier virtual-clock, the processor may need to sort the cell positions in a service queue every time it receives a cell. This may result in significant processing overhead when the queue is large.

2.1.6 Fast Buffer Reservation Control Scheme

The end-to-end protocols of virtual connections typically will operate on the basis of larger data units consisting of many ATM cells. If cell discarding is done

on a cell basis rather than on a burst basis, each discarded cell is likely to belong to a different virtual connection. Therefore it would be desirable for a congested network to react by discarding cells on the basis of connection bursts so that the cell losses will affect as few connections as possible. Turner proposed a fast buffer reservation scheme [21] that allocates network resources to traffic bursts of variable rate connections in order to preserve the integrity of the bursts. For each connection passing through a given link buffer, the network keeps track of the activity on the connection by associating a state machine with two states: idle and active. Upon reception of the start cell of a burst, a prespecified number of buffer slots in the link buffer will be allocated to the connection if they are available and the state machine enters the active state. The connection is guaranteed access to those buffer slots until it becomes inactive, which is signaled by a transition to the idle state, and then the buffer slots will be released. If the unused slots in the link buffer are less than the prespecifed number, the whole burst will be discarded. The virtual connection admission control makes acceptance decisions based on the estimated probability of the instantaneous demand for buffer slots exceeding the buffer's capacity, called the excess buffer demand probability, so that the burst discarding rate can be controlled to an acceptable level.

2.1.7 Congestion Control Schemes for Best-Effort traffic

With the admission control and bandwidth enforcement schemes that have been described above and a priority control mechanism functioning at the network element to support differential QOS, the traffic with predictable and relatively smooth characteristics such as Constant Bit Rate (CBR) and Variable Bit Rate (VBR) traffic can be handled efficiently and effectively. However, a large number of existing data networking applications such as LAN interconnection produce very bursty and highly unpredictable traffic and they are not suitable for strict admission control. In the past, window-based flow control mechanisms have been widely used for traffic control

in networks and have served well for reliable data transfer applications in low-speed network environments. With the realization of high-speed channels and the need to support diverse service requirements of new data applications, it has been found that the traditional sliding window control schemes tend to function slowly under such an environment and thus are incapable of providing effective control functions.

Several researchers have investigated new bandwidth management schemes that provide adequate performance improvement for high-speed data applications. One promising approach is to allocate some minimal resources to the class of traffic but allow additional traffic transmission on a best-effort basis with no performance guarantees. Two classes of closed-loop feedback control mechanisms have been proposed to regulate the traffic entry: rate-based [82] and credit-based [81] schemes. It has been shown that both schemes can be engineered to provide high resource efficiency in ATM environment. In addition to these works, an adaptive scheme where both flow window size and buffer size of a connection can be dynamically allocated during a call duration was proposed by Doshi and Heffes [18] for large file transfer applications.

2.2 Switching Technology and Control Strategies

Fast packet switches have been extensively studied in the literature because of the wide range of environments where they can be applied. An overview of the ATM switch architectures and products will be presented in Section 3.1. A good survey on commonly used switch architectures and discussion of issues pertaining to implementation and performance can also be found in [25].

Due to the unscheduled nature of arrivals to a packet switch, inevitable blocking may arise if two or more inputs attempt to transmit packets simultaneously to the same output. An essential issue in realizing the fast packet switching is how to smooth the statistical fluctuations in packet arrivals to the switch. Hluchyj and Karol

[26] have proposed four different approaches for providing the necessary buffering in the switch: these are input queueing, input smoothing, output queueing and completely shared buffering. Their study has shown that the system can achieve optimal throughput and delay performance if all the queuing is done at the output.

A detailed analysis and performance study of an output-buffering packet switch based on exponentially distributed packet length assumption were presented in [27]. Iliadis has demonstrated that under imbalance traffic the shared output buffering arrangement performs better than the dedicated output buffering arrangement [28]. However, a potential problem with the shared output buffering is that one heavily loaded output might monopolize use of the shared buffer, thereby adversely affecting the performance of other outputs. There are many studies devoted to finding the best approach to allocate a limited buffer space among outgoing links. A detailed comparative performance analysis of various buffer management schemes can be found in [29]. A major drawback that limits the application of these early works is the simplifying assumption of nonvarying traffic loads made in developing these schemes. Recently, Tipper and Sundareshan have presented an adaptive algorithm to dynamically allocate buffers under varying load conditions [30].

The service scheduling problem at a multiplexing or switching node has also drawn lots of attention. It is well recognized that priority queueing schemes can be used as an effective scheduling method to satisfy different delay requirements of mixed ATM traffic [31]. Several dynamic priority schemes have been proposed and studied for optimally controlling the performance tradeoff among multiple service classes [32, 33, 34, 35, 36]. Two dynamic priority schemes, Minimum Laxity Threshold (MLT) and Queue Length Threshold (QLT), were presented in [34]. With the priority controls, the priority queues are served in a dynamic manner so that the performance degradation for the low priority traffic can be reduced as the high priority traffic is overloaded. Their performance on scheduling two classes of traffic (realtime and

non-realtime traffic) was compared to the performance of two other conventional scheduling disciplines, first-come-first-serve and strict priority scheme. It has been shown that by employing the dynamic priority scheme, desired performance levels can be achieved for both the traffic classes if an appropriate threshold is chosen.

Lim and Kobza have proposed a Head-of-the-Line with Priority-Jumps (HOL-PJ) priority scheme to explicitly control the delay performance of different classes of delay-sensitive traffic [37]. Although the HOL-PJ control scheme provides delay guarantees for traffic with different service requirements, a potential drawback of this scheme is the significant processing overhead required for monitoring cells for time-out and moving cells to the next level priority queue. Lately, Oh, et al. have presented a dynamic priority method [36], in which the priority assignments are carried out based on the number of cells queued in each service class buffer and their waiting time.

There are some research efforts devoted to the analysis of applying priority discarding schemes as a local congestion control to satisfy diverse loss requirements of different traffic classes [38, 39, 40]. Some of the ATM applications such as voice communications can tolerate limited loss of information without significantly degrading service quality. The priority discarding strategies recognize the different cell loss requirements of traffic classes and discard information with bias. It has been shown that such a prioritized system is capable of achieving better performance than non-prioritized systems [38]. Furthermore, a priority discarding scheme can be used in conjunction with a special encoding technique so that the impact of cell discarding can be reduced to a minimum. Goodman [41] has proposed an embedded coding method to segregate the coded bits for a segment of speech into separate packets according to their importance to the decoding process. Priority is given to cells carrying more important information when network congestion occurs.

The cell loss probabilities of different loss priority classes subject to various space priority strategies for a finite buffering system were analyzed and compared in several research works. Kroner [73], Hebuterne and Gravey [75] and Doshi and Heffes [76] analyzed a *push out* scheme with different queueing models. In the control scheme, an arriving low priority cell is lost if the buffer is full, while an arriving high priority cell is allowed to push out a low priority cell in the queue in order to make room for itself if the buffer is full. Another space priority strategy that also has drawn lots of attention is the *partial buffer sharing* scheme, where a low priority cell is admitted to the system only if the current queue length is less than a pre-defined threshold; otherwise it is lost. Lin and Silvester [70], Kroner [73] and Meyer, et al. [74] have studied the cell loss probabilities of a binary-class queueing system employing the partial buffer sharing scheme. Petr and Frost [69] deal with the problem of choosing the optimal set of nested thresholds for a multiple-class system. A searching procedure is proposed in their study for the selection of optimal loss thresholds based on a simplified queueing model.

CHAPTER 3
PACKET SWITCH DESIGN FOR ATM NETWORKS

3.1 General ATM Switching Mechanism

Asynchronous Transfer Mode (ATM) which was initially proposed as a standard for the emerging BISDN is clearly the most appropriate transport technique to use for communication services with diverse traffic characteristics and service demands. However, while ATM certainly offers great flexibility in handling the wide diversity of bandwidth and latency requirements resulting from the integration of broadband and narrowband services, the timely implementation of ATM creates new challenges in switching technology. Over the past several years enormous efforts have been devoted to the research and development of ATM switching systems. In this chapter, a survey is given of the design and implementation of ATM packet switches and the relative advantages and disadvantages are discussed. This research has considered three major approaches to ATM switch design: Shared Medium, Shared Memory and Space Division. In addition, the Loral CPS-100 ATM-based switch is presented as an example of the fast packet switches based on shared medium/output buffering schemes. The rest of this section gives an overview of the ATM switching technology and products. Section 3.2 discusses the performance characteristics of different switch architectures. Section 3.3 describes the CPS-100 switch architecture and functional blocks in detail and Section 3.4 discusses the performance characteristics of the CPS-100 packet switch.

The number of manufacturers working on ATM switching technology has increased dramatically over recent years. Some of them are building small ATM

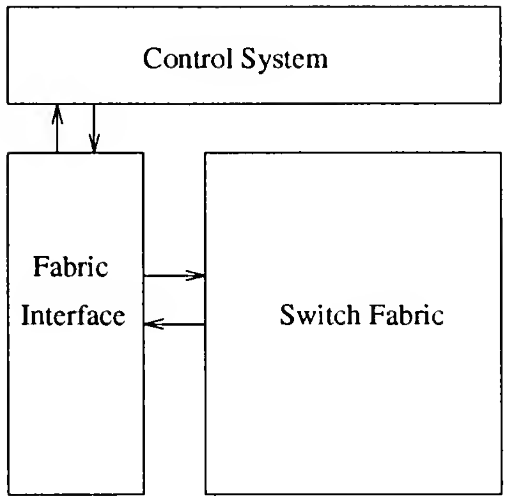


Figure 3.1: General ATM Switch Architecture

switches for LAN markets, and others focus on larger switches for MAN or WAN applications. An ATM switch may generally be divided into three major components: (1)*system control*, (2)*fabric interfaces* and (3)*switching fabric* [3, 5] (see Figure 3.1). System control is responsible for high level control functions such as virtual connection management, maintenance and administration, which are mainly performed by software. Fabric interfaces provide connections to the external facilities. The functions of a fabric interface may include speed and format conversion, synchronization, Virtual Connection Identifier (VCI) translation, signaling and routing information identification, load monitoring, congestion control and so on. The switching fabric performs the basic switching function based on the routing information added by the fabric interface to every incoming cell. Of these three components the switching fabric is the most significant part of an ATM switch since it decides the performance, complexity and cost of the system design [5].

3.1.1 Switching Fabric

The switching fabric may fall into one of the two forms: *time division* or *space division*, depending on its switching scheme [25]. The key distinction between these two kinds of switching fabric is the sequential versus parallel routing of cells. In time

division, all the input cells are multiplexed into a single common stream before being routed to their output destinations. The core of the time division switching fabric may be either a high-speed shared medium or a shared memory. Since every cell flows across the shared medium/memory, the medium/memory bandwidth should be sufficiently large to accommodate simultaneously all incoming traffic. The bandwidth thus places an upper limit on the total capacity that can be supported by the switch architecture. A number of time division switch designs with a total capacity up to a few Gbps have been developed and implemented, for example, Prelude [42], Hitachi [43] and Paris [44]. Although using today's state-of-the-art VLSI technology, it is possible to implement a time division switching fabric with a large switching bandwidth of more than a few Gbps, the switch design is still limited by other electrical and physical factors which could affect the information transfer, for instance, the reflections on the multiple stubs of the shared medium [45].

Some manufacturers [44, 46, 47, 48] have also developed shared medium/memory based switching elements as a chipset or a single chip so that large switches can be constructed in a modular way by interconnecting smaller building blocks. For example, the CMOS VLSI of the ATOM switch from NEC [47] was realized by using the Bit-slice technique for organizing the data and control portion of the switch elements and topology frame technology, which eliminates the need for backplane connections by allowing boards to be mounted edge-to-edge. In [46], a shared memory switching element which was realized by a single CMOS chip is described. It has been shown that, under the assumption of a combination of Bernoulli input traffic patterns and probabilistic routing, the cell loss probability can be less than 10^{-12} at 85% load per output of the switching element. The shared memory design has been proven able to improve significantly the cell loss probability of the switch, since all the input and the output queues can allocate the memory buffer dynamically on demand [5].

In space division, there exist parallel physical paths between the input ports and the output ports so that multiple cells may be transmitted through the switch concurrently. This makes the capacity of the space division switching fabric theoretically unlimited although, in practice, it is still restricted by other factors such as power consumption or the physical size of the switch. The concurrent transmission of multiple cells through the switching fabric also introduces the possibility of conflict in setting all required paths simultaneously. This phenomenon is usually referred to as internal blocking and is a significant factor which limits the throughput of the switch. The maximum throughput that can be achieved for a particular switch architecture and a given hardware complexity, in terms of the number of required stages and switching elements, are the central issues underlying space division switch design. Space division switches may be further classified according to a variety of aspects, for instance, single versus multiple path, (internally) buffered versus unbuffered, and blocking versus nonblocking. The Crossbar network [25], the Banyan network [49] and its variations, the Benes network [50] and the Clos network [47] are some well-known examples in the general family of space division switches.

3.1.2 Buffering

Buffering strategy is another important issue affecting the performance of all ATM switches. Due to the statistical nature of arrivals at a switch, a number of cells from different input ports may simultaneously be addressed to the same output port. At most one of these contending cells is allowed to pass through the switch immediately while the remaining ones must be queued for later transmission. This kind of contention, commonly referred to as output contention, is unavoidable and must be accommodated by some form of buffering. There are a number of options for buffering strategy [51]: internal buffering, input buffering, output buffering, as well as combinations of the above. The throughput of the switch is largely dictated by the type of buffering selected. For input buffering, the cell arriving at input k is stored

in a buffer corresponding to input k, while for output buffering, the cell destined for output k is stored in a buffer corresponding to output k. On the other hand, internal buffering provides buffers inside the switching fabric where contentions occur. It has been shown that a switch with output buffering can achieve optimal (unity) throughput performance while a simple input-buffering switch, due to head-of-line (HOL) blocking, is always restricted to a maximum possible throughput of less than unity [52]. It has been proven that, under the assumption of uniformly distributed traffic load, the maximum possible throughput of an input-buffering switch is 0.586 when N (the number of incoming/outgoing links) is very large. The throughput performance of input-buffering switches can be improved by the use of more sophisticated control mechanisms. A parallel "ATOM" (ATM Output-buffering Modular) switch architecture was proposed in [54]. With this architecture, a 32 by 32 ATM switch with a capacity of 9.6 Gpbs was realized using 70 Mbps BiCMOS technology and a simple resequencing technique was used to keep the sequence integrity for cells sent out from the parallel switching planes.

3.2 Performance Study

A particular switch architecture may be evaluated from the viewpoint of a combination of important performance characteristics and architecture parameters, such as switch capacity, growth potential, hardware complexity, reliability, functionality, modularity, controllability and so on. It is difficult practically to single out any one among the architectures as the best. In this section the advantages and limitations of time division and space division switch architectures are discussed from several important aspects.

- *Capacity*: The time division approach is a feasible and flexible technique to implement small-size switches with a capacity up to a few Gbps. This class of architecture provides an efficient and flexible way to multiplex port interfaces

with widely differing access rates and protocols. With present levels of hardware and software sophistication, time division switches are less complex to build and are highly reliable. However, the basic time division switching architecture is not suitable for large switch designs. The entire switch capacity is limited by the bandwidth of the shared medium/memory, which restricts the number of ports that can be supported to a relatively small number. Thus for the construction of large switches such as ones for broadband central offices, a multistage or space division switch architecture must be adopted to achieve a larger capacity. Figure 3.2 indicates some examples of the potential application areas for time division and space division switch architectures.

- *Growth potential*: Excellent scalability is one of the attractive features of space division switches. Space division switches are generally constructed by interconnecting a number of fundamental switching blocks. Their capacities may be increased significantly by simply providing additional paths that can operate concurrently for transmitting cells. Thus this class of architecture can readily be scaled to large switch designs. On the other hand, the growth potential of time division switches is relatively poor since their capacities cannot grow beyond an upper limit.

- *Complexity*: Most of the space division switch designs have the important features of self-routing and distributed control, which allow each fundamental switching block in the switching fabric to make a very fast routing decision. However, they are generally more complex to construct and it is more difficult to analyze the performance characteristics than for time division switches. Space division switch designs require sophisticated fault detection and isolation procedures to enhance their reliability. Moreover, as the switch size increases,

the synchronization for data and control signals, VLSI integration, and interconnection structures become significantly more complicated.

- *Functionality*: Priority control and multicast operations are two functions that may be desirable in many applications, such as video conferencing and distribution services. In time division switches, these functions can be easily supported by modifying the buffer or memory read/write control circuit. These operations can also be achieved in space division type switches although it may require much more design effort and hardware implementation.

3.3 A Shared Medium/Output Buffering ATM Packet Switch Design

3.3.1 CPS-100 Switch Design Characteristics

The CPS-100 switch developed by Loral is an ATM-based switching system that supports ATM and various data services such as Switched Multi-megabit Data Service (SMDS) and Frame Relay for public networks. In the fast packet switching terminology described above, CPS-100 switch architecture can be viewed as a *time division, shared medium* switch fabric with interconnecting *time division, shared memory* switching elements. The switch structure is generally composed of the following four major components:

1. The *switching fabric* performs the basic switching function of transferring packets from input to output. The switching fabric in the CPS-100 switch is composed of a high-speed Virtual ATM (VATM) backplane bus, which is shared between all interface modules.

2. The *access interface module* provides an interface for customers to access the services supported by a network. The Customer Premises Equipment (CPE) attaches to an access interface module for transmitting and receiving data via a dedicated link.

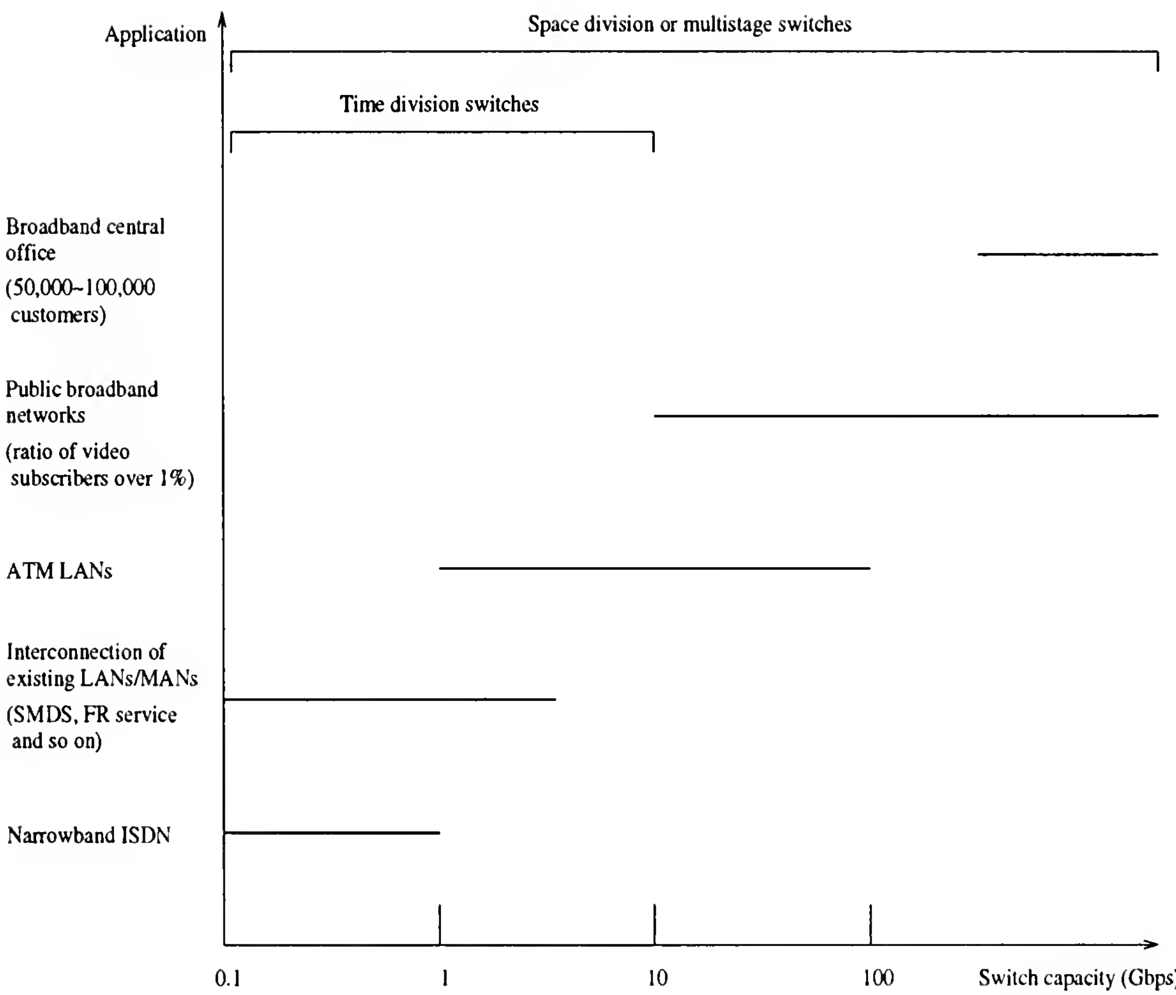


Figure 3.2: Potential application areas for time division and space division switch architectures

3. The *trunk interface module* provides an interface for interconnecting switching systems to support inter-switch communications.

4. The *CPU module* communicates with all the interface modules in the switch and performs the high-level control functions such as connection establishment and release, memory administration, bandwidth allocation and maintenance.

3.3.2 Interface Modules

Figure 3.3 illustrates the functional model of a simplified ATM network implemented with CPS-100 switches. Note that an access/trunk interface module can support multiple bidirectional links. External trunks or CPE links are all connected to the VATM via the access/trunk interface module, which consists principally of two controllers, namely the *ingress adaptor* and the *egress adaptor*. The *ingress adaptor* receives cells from the incoming link, buffers them as necessary and sends them through the switching fabric. Similarly the *egress adaptor* accepts cells from the switching fabric, provides any necessary buffering, and transmits the cells over the outgoing link. These adaptors also perform all the cell-by-cell processing functions such as Virtual Connection Identifier (VCI) translation, priority assignment and routing. Since the VCI is generally only local to each switch port, the VCI of each cell needs to be translated to the value assigned for the succeeding links. This operation is performed by a VCI translation table.

The ingress adaptor appends routing overhead on an incoming cell to specify the output link associated with the virtual connection to which the cell belongs. The ingress adaptor also looks up other connection type information such as the loss priority and the service priority of the cell to assist the egress adaptor in making its buffering and service scheduling decisions. Since all the cells of a virtual connection will follow the same route in traversing a switch, the sequence of cells is preserved, without the need for explicit sequence numbers.

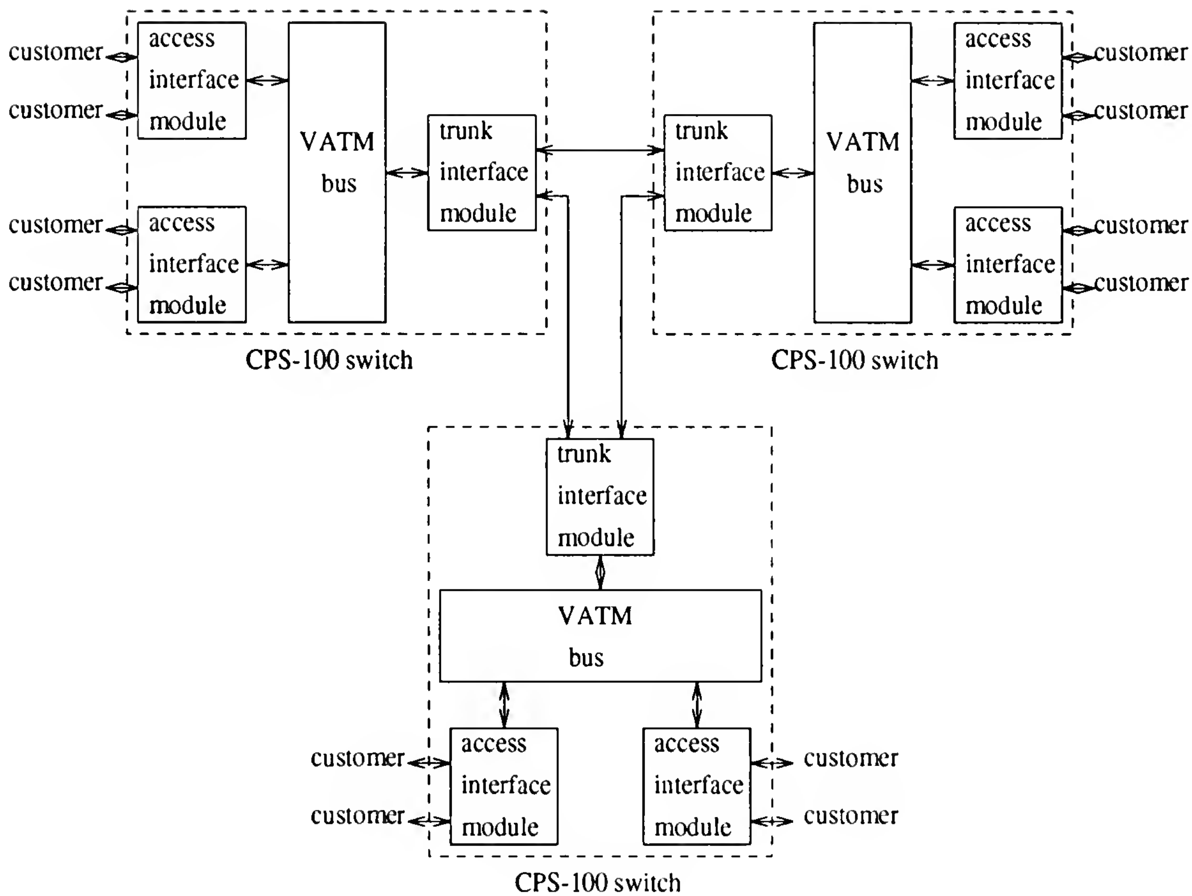


Figure 3.3: Functional model of a simplified ATM network

The most important function of the egress adaptor is to buffer temporarily those cells which cannot be immediately sent out. A finite buffer memory is implemented in the egress adaptor to resolve output contention. If congestion causes the buffers to be filled up, the controller has no choice but to discard cells. The buffer thus forms a critical resource that needs to be adequately controlled. A modified Partial Buffer Sharing scheme is implemented in the egress adaptor to ensure satisfactory loss rates for different loss priority classes and fairness in sharing the resource among competing users. Moreover, a priority-dependent service scheduling scheme is implemented to control the order in which cells are sent from the buffer to the output link. The service scheduling control is necessary for meeting the different switching delay requirements for different service priority classes. Service and loss priorities are assigned to each virtual connection during connection establishment based on the delay and loss requirements specified for the particular connection.

Figure 3.4 shows the schematic of the ingress and egress adaptors for a two-link interface module. The functions of the components indicated in the figure are summarized as follows.

- The FIFOs are small First-In-First-Out queues which store cells to accommodate short data bursts which exceed the processing speed.

- The Ingress Message Processor is responsible for looking up the routing and service information for incoming cells. It searches a routing table to find out the output port address to which a cell needs to be switched and attaches the routing overhead to the cell, which is then sent to a FIFO queue. The Ingress Message Processor also support multicast operation by duplicating the incoming multicast cells and appending appropriate output port address to each of the copies.

- The VATM bus performs the switching operation of transporting cells from an ingress adaptor to an egress adaptor according to the routing information. The high-speed bus operates synchronously so that one cell can be transmitted across the bus during each transmission cycle.

- The function of the one-cell buffer in the egress direction is to provide timing independence between components. The FIFO connected to it is sensitive to the backpressure from the one-cell buffer, that is, the FIFO will only send a cell to the associated one-cell buffer as the previous cell has left the buffer and served by the Egress Message Processor.

- The Egress Message Processor controls the egress buffer space usage. A modified Partial Buffer Sharing scheme is activated upon receiving a cell to determine if the cell can be accepted to the Buffer Memory allocated for its output port address.

- The Egress Buffer Memory stores cells while they are waiting to be serviced by the Egress Protocol Processor. The buffer memory of each port in an interface module can remain dedicated or they can be merged to achieve a lower probability of memory overflow.

- The Egress Protocol Processor extracts cells sequentially from the buffer memory by referring to their service priorities and then removes the routing overhead and assigns new parameters for the outgoing cell according to the communication protocol.

The throughput of each individual component is dependent on the link speed of the external link. The throughput of the Ingress Message Processor is designed to be marginally faster than the link speed (e.g. 1.1×Link Speed). A relatively small buffer in the ingress adaptor is also provided. Moreover, the Egress Message Processor

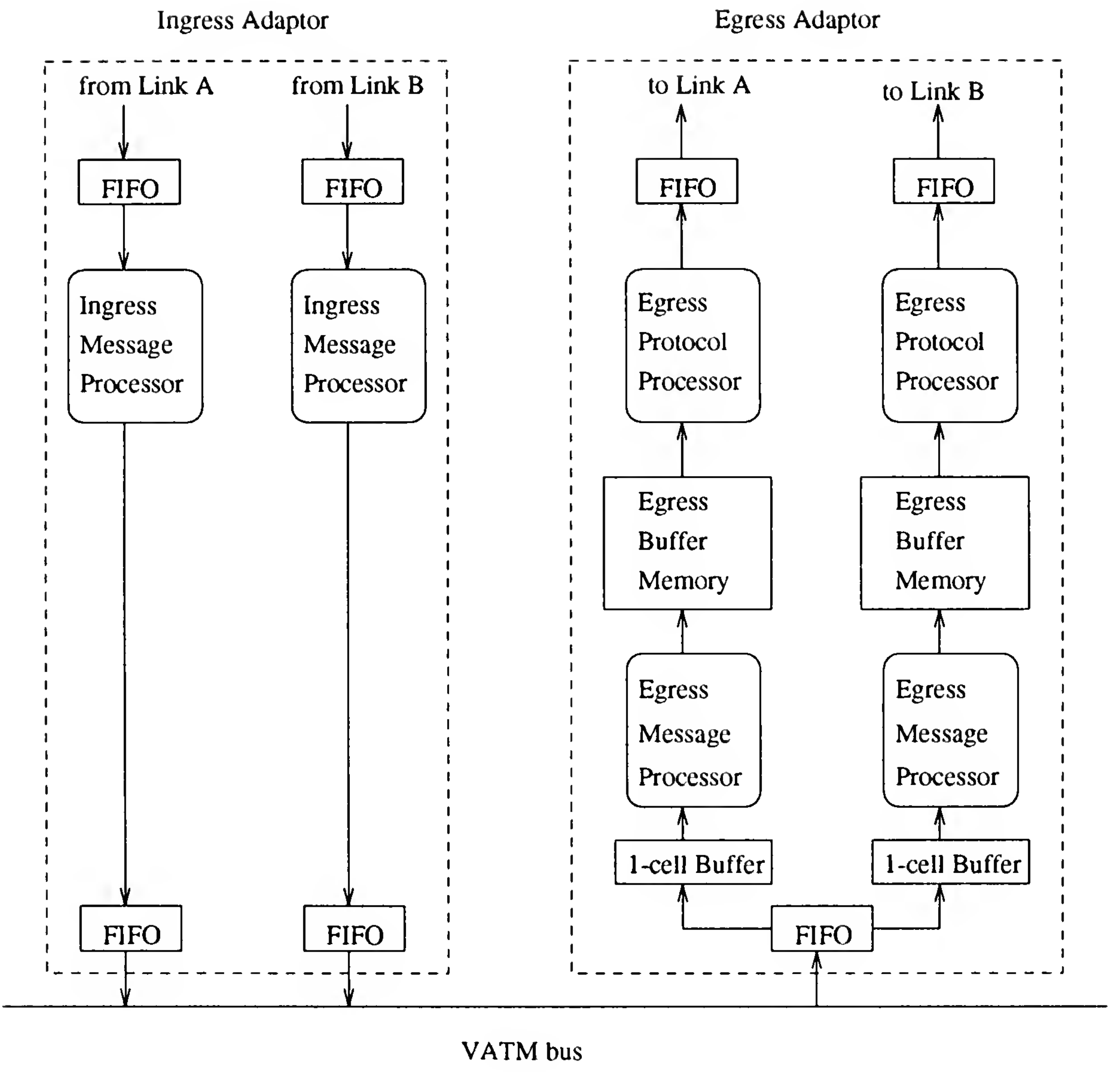


Figure 3.4: Schematic of a two-link access/trunk interface module

in the egress direction is capable of functioning at a comparable speed as the VATM bus so that the high-speed cell stream from the VATM bus can be accommodated.

Upon receipt of an incoming cell, the ingress adaptor will seek access to the VATM bus. Only one ingress adaptor can have access to the high-speed bus at any time. A bus arbitration scheme is implemented for resolving contention for access among different ingress adaptors. However, since the shared medium operates at a substantially higher speed (approximately 5 Gbps) than the link interface modules do, the CPS-100 can be seen as a non-blocking switch, having no internal buffering and with no contention visible from outside the switching fabric. The high speed of the switching fabric reduces input contention and results in major queueing taking place at the egress adaptor. Thus the switch architecture is expected to give performance similar to an output-buffering switch and, therefore, takes advantage of the optimal delay/throughput performance that can be provided by an output-buffering switch [52].

3.3.3 Simulation Approach

The simulators which accurately reflects the hardware design of Loral's CPS-100 switch have been developed using two different simulation packages, the OPtimized Network Engineering Tools (OPNET)[1] and the General Purpose Simulation System (GPSS)[2]. However, we found that the GPSS is incapable of supporting completely the simulation works due to memory constraint and flexibility of the software. Thus the majority of our simulation studies were performed by using the OPNET package.

The OPNET tool is an event-driven simulation package designed specifically for the development and analysis of communication networks. It offers a hierarchical modeling approach and graphical interfaces for users to model network and system

[1]The tool OPNET was developed by MIL 3, Inc.
[2]The GPSS was developed by Minuteman Software

specifications. The OPNET package provides a detailed and flexible modeling environment with a significant reduction in the extensive software development process which is typically associated with complex system modeling.

Each hardware functional block in the switch is simulated with a parameterized module, and modules communicate through packet flow and control signal connections. Each module is associated with a process model which contains the logic of model operations. The processing delay of functional blocks, traffic control process and switch architecture can be modified with the assignment of different parameters, process models or interconnection structures. An arbitrary network configuration can be constructed easily by interconnecting a number of switch simulators and traffic sources that model the specified traffic patterns.

The switch simulator and its output have been verified by Loral's and they contribute to the development of the traffic control schemes as well as the switch design. The simulator also serves as an experimental platform for the feasibility and performance evaluation of the traffic control mechanisms that are considered in this dissertation.

3.3.4 Simulation Results

During the research period, we have performed numerous simulations and analyses to identify the switch characteristics with different traffic models and evaluate implementation complexity as well as effectiveness of various traffic control schemes. In this section, part of the results of our simulation study are presented to demonstrate the basic CPS-100 switch characteristics. We show that the network delay performance can be controlled effectively by implementing different scheduling schemes in the switch.

The simulation was performed based on a network configuration as shown in Figure 3.5. In the network configuration, two CPS-100 switches are interconnected to each other via a trunk link. On each side of the network, there are a number of

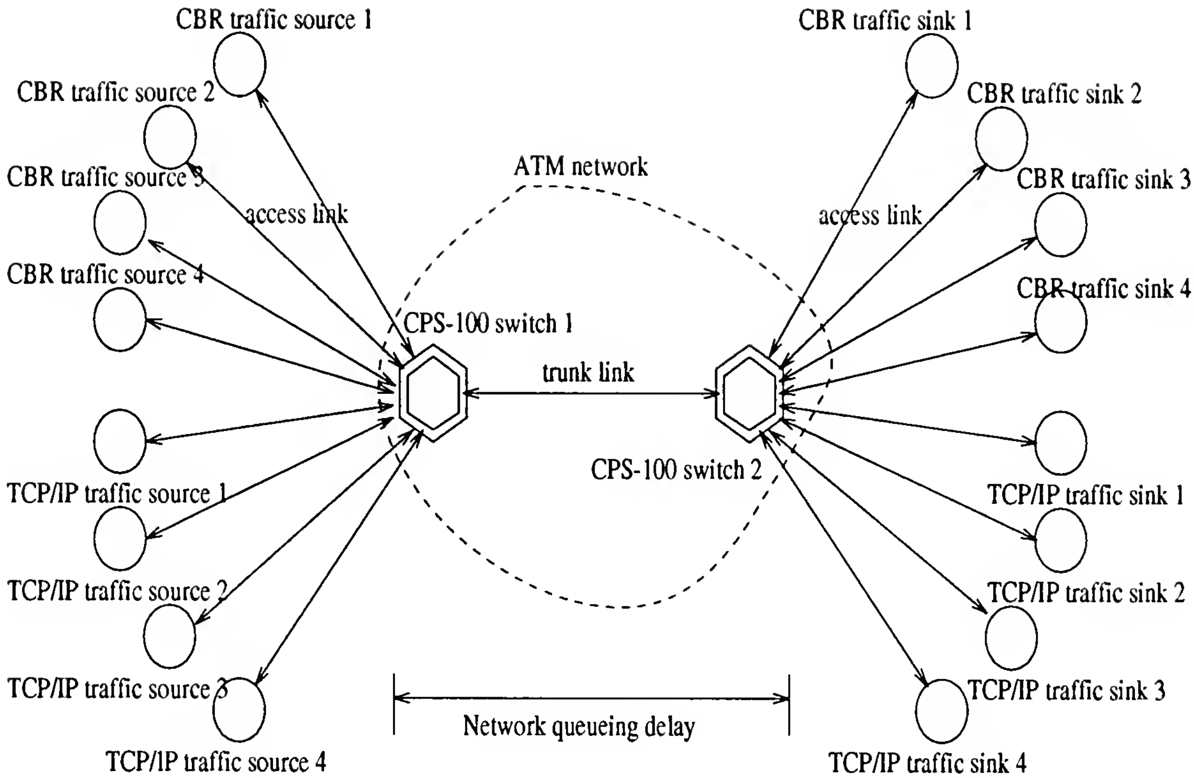


Figure 3.5: The network configuration for simulation

customer premises equipment (CPE) connected to the switch. It is assumed that each CPE on the left side makes a connection to the CPE on the right side and simultaneously transmits data through the trunk link. Since all the traffic aggregates on the outgoing trunk port of switch 1, the egress buffer memory and the outgoing link bandwidth of the port become performance bottlenecks, where congestion may occur when too many traffic bursts arrive in a short period.

The assumptions made for the input traffic loading are summarized as follows. Suppose there are a total of eight virtual connections in the network sending traffic to the CPEs on the other side. The traffic sources of virtual connections #1 to #4 are Constant Bit Rate (CBR) sources which generate constant-rate cell streams. The CBR traffic is sensitive to the queueing delay in the network and needs to handle in a higher priority than other traffic. On the other hand, the traffic pattern of virtual connections #5 to #8 represent bursty data sources which occasionally produce long

cell bursts at the peak link rate. A TCP/IP connection with a 64-kbyte adaptive flow window and a 8-kbyte packet segment is simulated for each data source.

Assume that each virtual connection is sequentially assigned with a different scheduling priority: virtual connection #1 has the highest scheduling priority, say 1, and virtual connection #8 has the lowest scheduling priority, say 8. The CBR traffic (i.e., cells with scheduling priority 1 to 4) is served by the scheduling server of the outgoing trunk port in an absolute priority method, by which cells with higher priority are always processed in advance, as long as there is one queued in the buffer. On the other hand, the data traffic is treated in a relative priority method which is similar to the weighted fair queueing scheme studied in [20]. The available link bandwidth left after serving the CBR traffic is allocated between different scheduling priorities in a weighted manner. The higher the scheduling priority, the larger its weight. For instance, if the scheduling priority 5, 6, 7, and 8 has a weight of 4, 3, 2, and 1, respectively, the scheduling server will divide its bandwidth in serving the different queues proportionally to the weighting (assume that each scheduling priority queue is never empty). The unused bandwidth of higher priority traffic will be utilized by the incoming cells with lower priority.

For simplicity it is assumed that the outgoing trunk port has sufficient buffer space to accommodate all incoming traffic. For bursty data sources, the duration of the idle time between traffic bursts is assumed to be exponentially distributed. The traffic distribution between traffic sources is uniform, that is, each traffic source contributes about 12.5% to the total trunk load. The link speed of all the links in the network is T3 (45 Mbits/s) and the propagation delay on the links is assumed to be negligible. The weights for scheduling priority 5, 6, 7, and 8 are assumed to be 4, 3, 2, and 1, respectively. Twenty seconds of network time is simulated.

The mean and maximum network queueing delays for each scheduling priority are studied. The network queueing delay is defined as the time interval between a

cell arriving at an ingress adaptor of the first switch to the cell leaving the egress adaptor connected to its destination of the second switch (see Figure 3.5). It is observed that the queueing delay within the egress buffer at the outgoing trunk port dominates the network queueing delay experienced by a cell because the port is a congested point in the network configuration. Figures 3.6, 3.7 and 3.8 show the mean, maximum and variance of the network queueing delay for each scheduling priority as a function of the trunk link utilization. It is seen that the statistics of the network queueing delay of CBR traffic remain nearly a constant as traffic load increases. This is mainly because the cell interarrival time of a CBR connection is a constant. Since the cells arrive at the buffer periodically and the scheduling server will try to process these cells as soon as they come in, these cells will always have a very small queueing delay even when the traffic load is high. This result also implies that with an appropriate traffic shaping function exercised at the network entrance point to smooth the traffic pattern and the selection of an adequate service discipline, a maximal network queueing delay and cell delay variation can be guaranteed. On the other hand, the network queueing delay of the data traffic increases exponentially as the trunk link utilization increased. The lower the weight of the scheduling priority, the larger the increment of the queueing delay. With the assignment of appropriate weighting numbers, the mean queueing delay of the scheduling priority classes can be controlled within a range.

In addition, we study the time average and maximum egress buffer utilizations of the outgoing trunk port for each scheduling priority (see Figures 3.9 and 3.10). Again the trunk link utilization has very little effect on the egress buffer utilizations of the CBR traffic. In general, the time average egress buffer utilizations are proportional to the mean queueing delay for every scheduling priority. Figure 3.10 shows the maximum instantaneous egress buffer utilizations for each scheduling priority. These numbers are the maximum buffer capacity required so that no cell were

dropped over the simulation interval. Note that the maximum buffer utilizations of the data traffic, which are randomly distributed within a range, can go much higher than the corresponding time average buffer utilizations. The results shows that the instantaneous buffer utilization is closely related to the burstiness of the incoming traffic and the service discipline. For CBR or other relatively smooth traffic, a small buffer is sufficient to cope with the traffic variation. On the other hand, for the highly bursty data traffic, the occurrence of buffer overflows is almost inevitable for a finite buffer system unless additional flow control schemes (such as the schemes that are discussed in Chapter 6) are applied to the network.

3.4 Advantages and Limitations of CPS-100 Switch

A principal advantage of the CPS-100 ATM switch structure is its flexibility to support multiple kinds of port interfaces. It has been widely recognized that ATM technology is targeted to support a rich mixture of broadband and narrowband services, as well as the high-capacity interconnection of existing data networking applications. Thus it is important for a switching node to provide different kinds of interfaces for the diverse applications in a highly flexible and expandable way. In the CPS-100 switch, this is achieved through a modular design that places the ports on separate interface modules. Various combinations of port interfaces, link access speeds and physical media can be supported easily by implementing different interface modules on the switch. A CPS-100 switch can be configured to support up to 16 of these interface modules.

Another advantage of the CPS-100 switch is that with this architecture dynamic priority functions and multicast operations can be supported flexibly with no additional hardware complexity. This function is preferable since some applications such as ATM LANs require the ability to support multiple guaranteed classes of services and full-bandwidth multicasting. Each interface module can implement suitable

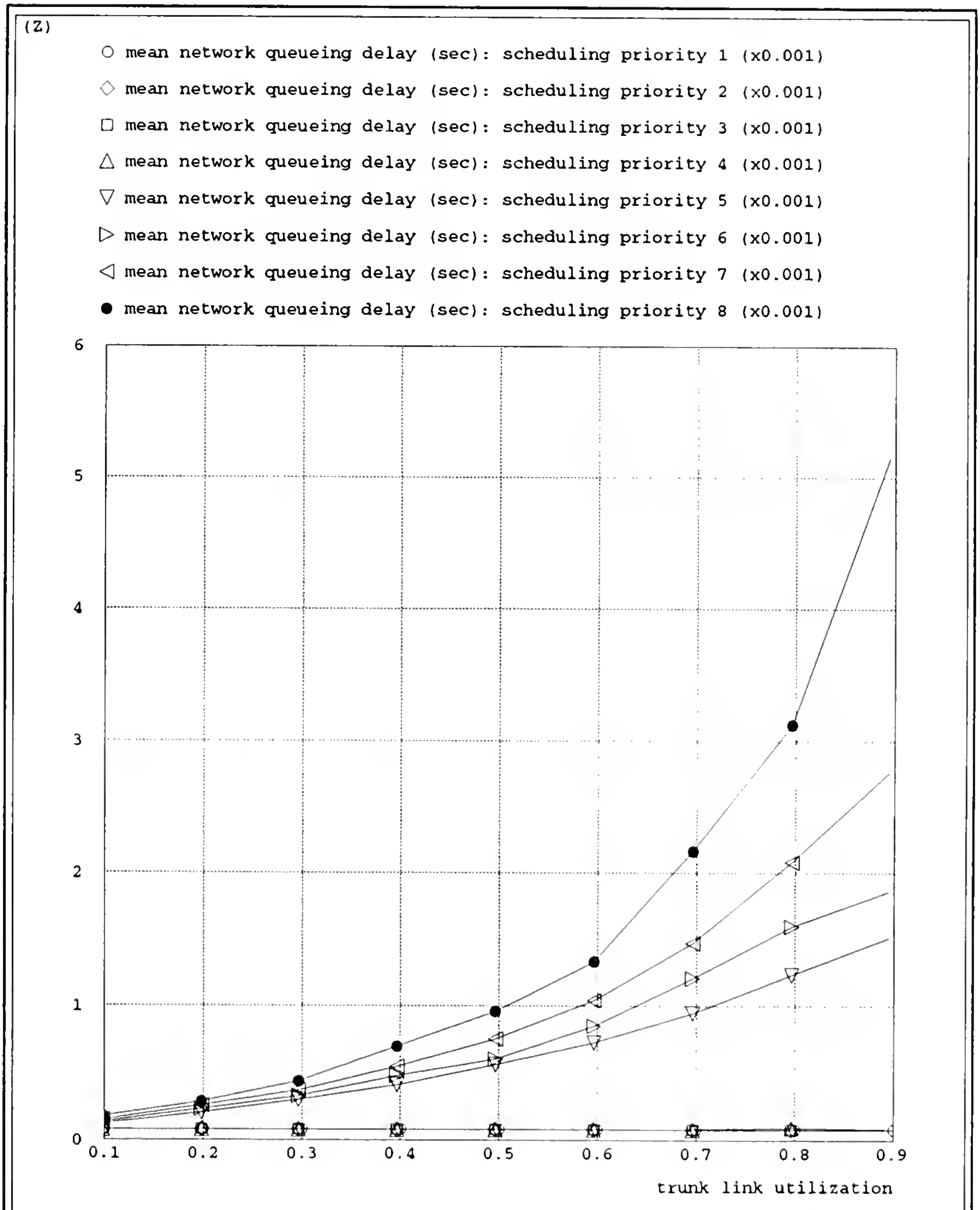


Figure 3.6: Mean network queueing delay of different scheduling priorities

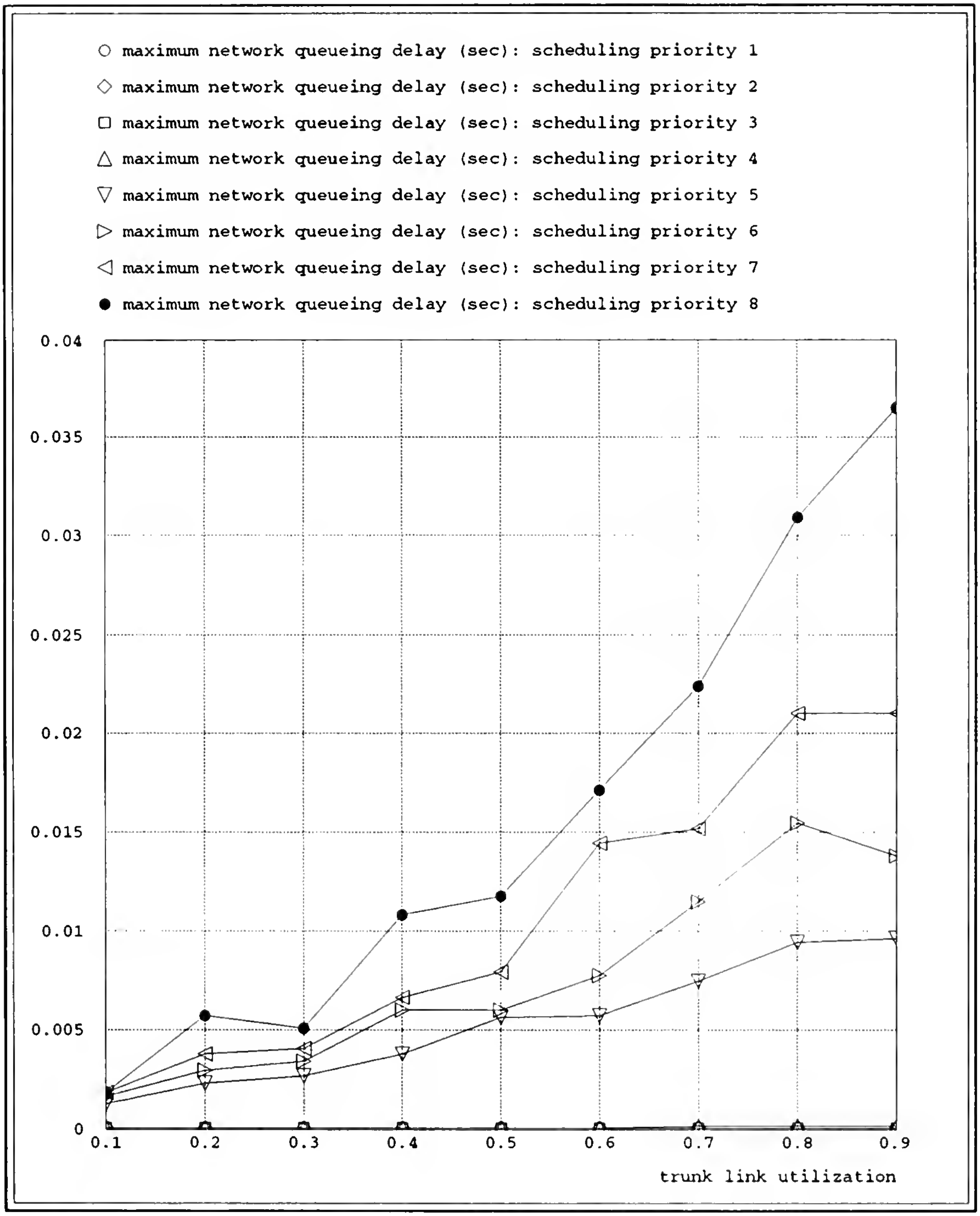


Figure 3.7: Maximum network queueing delay of different scheduling priorities

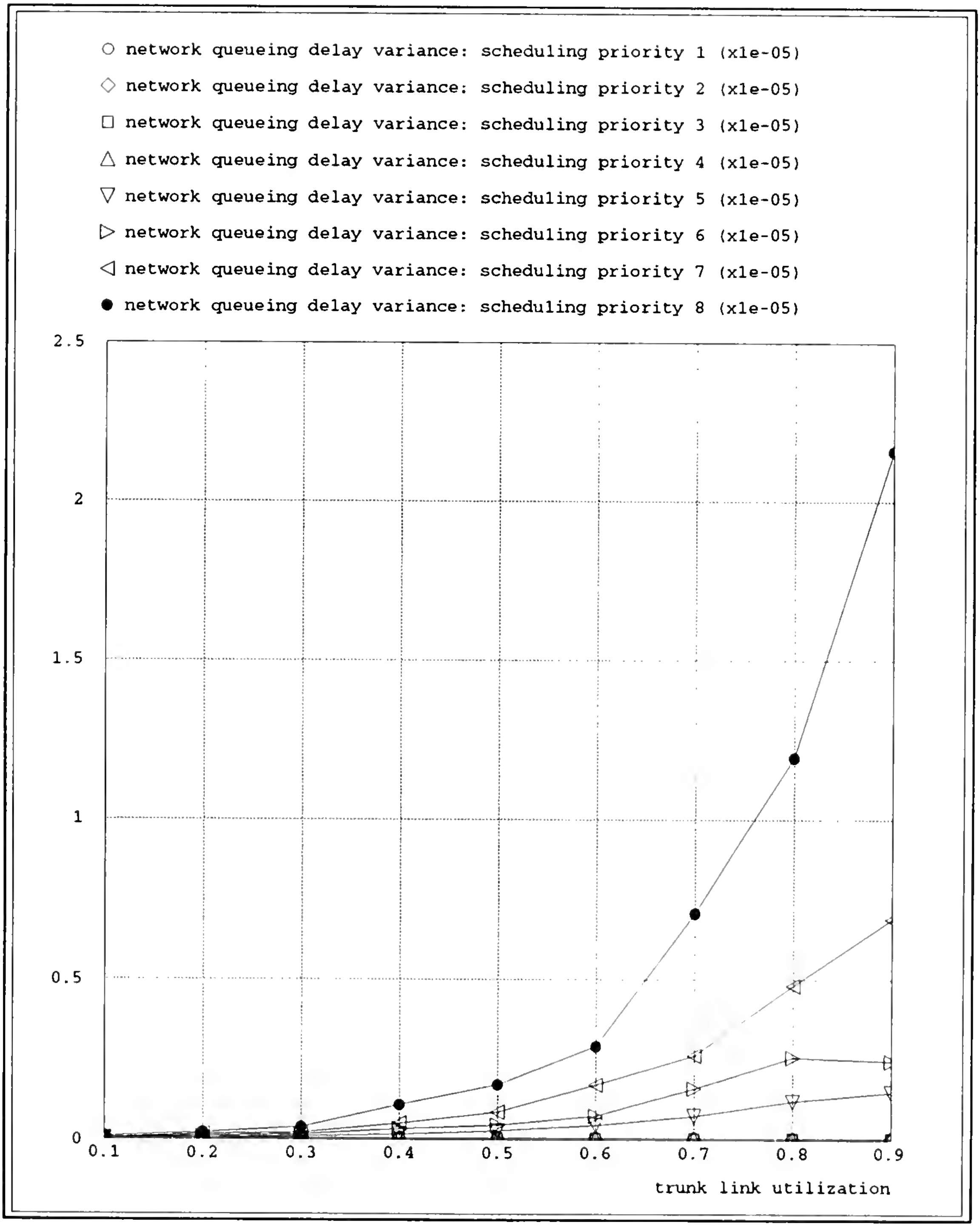


Figure 3.8: Network queueing delay variance of different scheduling priorities

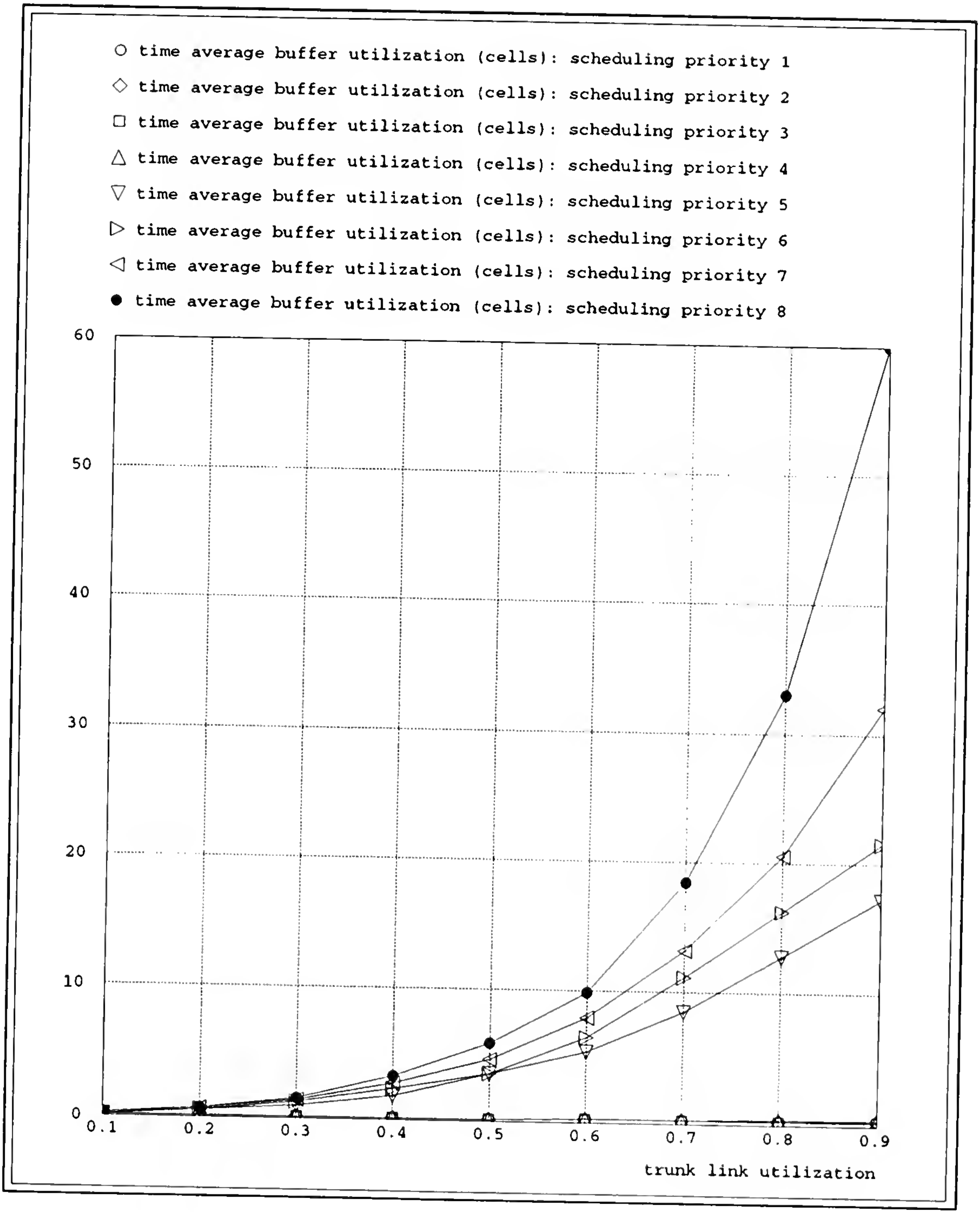


Figure 3.9: Time average egress buffer utilization of different scheduling priorities

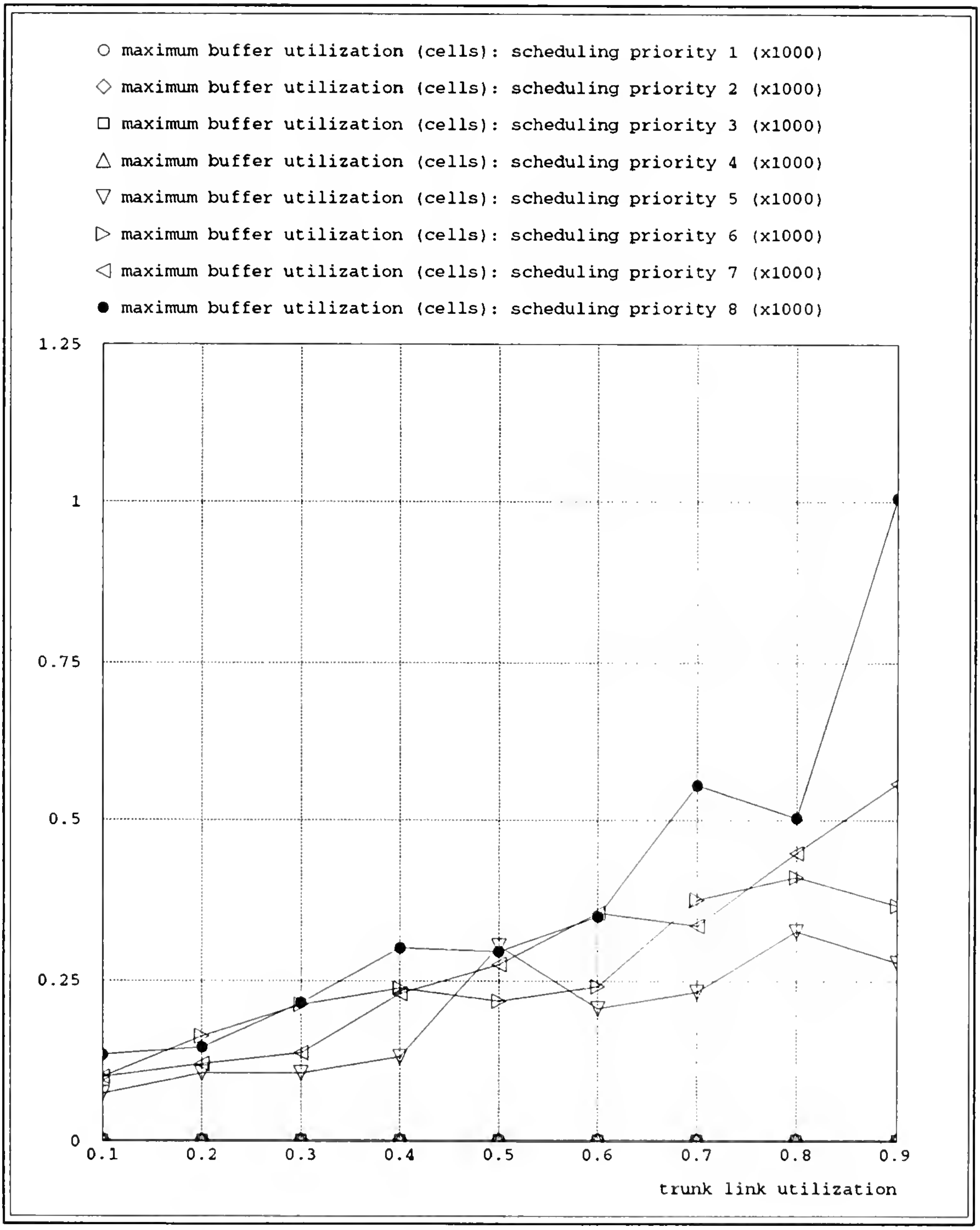


Figure 3.10: Maximum egress buffer utilization of different scheduling priorities

buffer management and scheduling policies based on the traffic characteristic as well as loading conditions to best serve the delay and loss requirements of different traffic classes. In addition, since the major queueing takes place at the output buffer, each interface module can perform load monitoring on the buffer usages of different traffic classes to support congestion control operation. Additionally the buffer space within an interface module can be allocated dynamically to each port, so as to achieve a lower probability of memory overflow.

On the other hand, as with other shared medium/output buffering ATM switches, the CPS-100 switch architecture has the typical limitation of requiring the high speed switching fabric to accommodate simultaneous arrivals from different input ports. The switching fabric and output buffer must effectively operate at a much higher speed than that of each interface port. The implementation of the high-speed bus and buffer memory could be complex if the required memory access speed is very high, thus limiting the capability of the switch to support very high speed interface ports.

CHAPTER 4
AN INTEGRATED ATM TRAFFIC CONTROL FRAMEWORK

4.1 ATM Traffic Control Requirements

The accommodation of the large spectrum of applications implies that ATM services must be differentiated based on QOS offered to the users [1]. The traffic characteristics and QOS requirements of different applications of ATM networks are likely to be very diverse. A successful deployment of ATM technologies will require a robust and flexible traffic management strategy to engineer network resource efficiently and support diverse service requirements. The traffic management schemes which are adopted by standard bodies will not only affect directly the complexity of network system hardware but also determine if ATM applications can be supported economically.

In this chapter, the initial service categories that will be offered on an ATM network are addressed and the necessary traffic management functions are studied in detail. We propose an integration of these traffic management schemes to provide a complete control framework to manage network congestion preventively as well as reactively and allow cooperative control actions at distributed network systems.

As mentioned in Chapter 1, traffic sources usually have traffic states that can be characterized in call level, burst level, and cell level. Therefore network congestion should be evaluated and controlled at different levels and time scales. Congestion can happen at call level when logical channels request resources (link and switching bandwidth, buffer space, etc.) that cannot be supported, when too many connections, which are admitted by the network to achieve high statistical gain, are active concurrently, or when a certain call exceeds its reserved bandwidth limits. It

can occur at burst level when too many long bursts of cells arrive and queue at some buffer causing buffer overflow. Congestion can also occur at cell level due to poor buffer management or as a result of upper layer operations such as retransmitting a complete data frame every time a cell loss occurs.

From a general point of view, the network should be engineered with sufficient resources to provide an acceptable level of connection blocking performance and allow some degree of future traffic growth [1]. *Network resource engineering* is one of the important traffic management functions which provides a long term control over the total amount of available bandwidth for all types of ATM services. Its major control function is to allocate sufficient resources for different service categories based on some understanding of predicted subscriber loads and usage characteristics.

When a call request is received by a network, the network control will determine if the required resources are available. If there are not sufficient resources for the connection, the network control will block the request or negotiate with the user for a less stringent service requirement. *Connection Admission Control* operating at the call level evaluates the impact of a new connection and allocates network resources to an accepted call on the basis of the bandwidth requested by the source [55]. Once the connection is established, the network has the responsibility to maintain the necessary resources for an acceptable QOS and monitor the source traffic to ensure that it complies with its declared parameters.

One way to alleviate burst level congestion is to implement large buffers in nodes to cope with traffic bursts [56]. With large buffers to store excess data during periods of burst level congestion, the achievable link utilization can also be improved. However, this approach will introduce added cell delay due to buffering of the excess traffic and increase the complexity of resource allocation.

Some applications, such as large file or image retrieval, produce single long bursts that can be treated as a single connection with individual establishment and

release phases. The *Fast Reservation Protocol* [56] operating at the burst level is a burst admission control method that allows in-call bandwidth negotiation when the traffic source needs to transmit a burst. Before a burst is transmitted, a traffic source sends a reservation request to each node on the connection path to reserve the required bandwidth. If there is not sufficient bandwidth available, a negative response will be returned immediately and the transmission of the burst will be blocked or delayed.

Due to the unpredictable nature of data applications, such as LAN interconnection, it is difficult to characterize the variability of traffic rate at connection setup time [57]. Allocating bandwidth at their peak cell rates results in low resource utilization and thus is inefficient for the burst-type data traffic. Most data applications can tolerate some delays but cell loss in the network will trigger retransmission to recover from information loss. A complete data frame has to be retransmitted each time cell loss occurs, regardless of severity of the cell loss. This traffic behavior is likely to cause sustained cell level congestion and lead to low packet *goodput* (the throughput of successfully transmitted packets). For these applications, a *feedback* control mechanism which can be invoked quickly to reduce the probability of packet retransmission should be provided.

Credit-based and *rate-based* flow control are two classes of control schemes that can regulate the volume of traffic admitted to the network according to the status of network load [56, 58]. Both kinds of the control schemes have been proposed to the ATM Forum for best-effort service and are currently under discussion. These control schemes provide a closed loop control to prevent or reduce the effects of network congestion and to optimize network resources. Such control approaches are attractive to the applications which may generate unpredictable traffic bursts, since the users need not specify their average or sustainable cell rates in advance of transmission. Moreover, with the flow control, traffic sources contributing to congestion can be

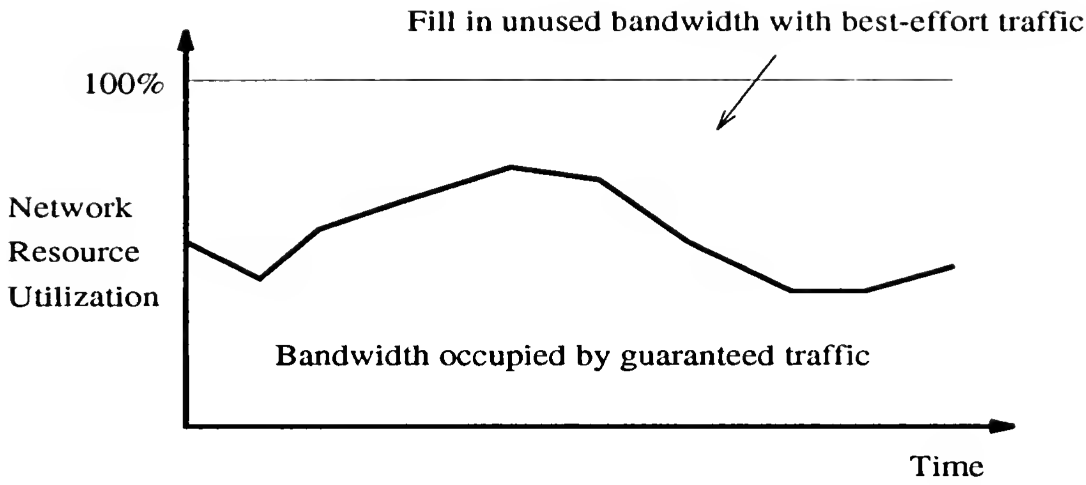


Figure 4.1: Network resource provisioned for guaranteed traffic and best-effort traffic

forced to reduce the speed at which they are sending. Thus if a connection admission control is used to support QOS requirements of guaranteed services, the bandwidth left over can be utilized efficiently by filling in with the best-effort traffic, as shown in Figure 4.1.

Once a new connection is admitted by network control, a monitoring mechanism is required to ensure that, during the information transfer phase, traffic submitted to the network does not exceed the traffic agreement established at the connection setup. This control function is performed by the *Usage Parameter Control.* The traffic policing is necessary to guarantee that violations by some users will not cause performance degradation for other well-behaved users.

Another important control mechanism that provides an effective cell level control and supports different QOS is *Priority Control.* When a new connection is established, a connection priority should be assigned to it based on the service requirement. The cells belonging to connections that need guaranteed service will be handled with higher priority by network nodes. In addition to the connection level priority, a cell loss priority (CLP) may be assigned to cells of the same connection to identify their significance or to mark the traffic that violates declared parameters.

Cells with lower cell loss priority (i.e., CLP bit set to 1) will be discarded first in case of resource congestion.

4.2 ATM Service Characteristics

To support different QOS requirements of users, three basic ATM service categories are identified corresponding to distinct user demands. In this section, the three basic ATM service categories that will be offered on an ATM network at the initial phase and their traffic characteristics are discussed.

4.2.1 Constant Bit Rate (CBR) Service

The Constant Bit Rate (CBR) type of service is intended for applications that will generate a constant or nearly constant cell stream, such as voice trunk and traditional video channels. This class of traffic has the most demanding service requirements. It is sensitive to delay and cell delay variation and typically cell loss in the network cannot be recovered by packet retransmission. This service type requires bandwidth allocation at connection's peak rate and service agreement sustained on a per-connection basis. Transmission of the traffic should be handled with highest priority in network nodes to ensure the specified QOS is guaranteed.

With the required network resources negotiated and allocated prior to transmission and maintained on a per-connection basis during the information transfer phase, ATM networks are able to provide strict QOS control for connections. As long as users conform to their traffic contracts and adequate resource protection is provided, cell level congestion will no longer be an issue. For this type of traffic, only a small buffer at switching nodes or multiplexing points is required to accommodate the effects of cell delay variation.

4.2.2 Variable Bit Rate (VBR) Service

The Variable Bit Rate (VBR) service offers a committed cell loss rate and delay for the users that require guaranteed service. This service is suitable for users or applications that generate traffic at a variable rate but have predictable statistical characteristics, such as video conference applications and distribution video. This type of traffic is typically delay-sensitive and with less stringent loss requirement. Recovery from cell error or loss is by concealment or cell substitution or by forward error correction, not by packet retransmission.

A gain in resource efficiency is achieved by performing statistical multiplexing of VBR sources, that is, by overbooking link capacity to allow bandwidth sharing between VBR connections. Link bandwidth will be allocated to the connections by taking account of their peak rates and some other traffic parameters defined for the service. With well-described statistical characteristics of traffic, it is possible to control congestion and cell loss below some probability by applying only admission control at connection setup. However, a larger buffer (about 1000 cells) at multiplexing points will be required to cope with burst level congestion and give better link utilization [58].

4.2.3 Available Bit Rate (ABR) Service

Currently, the increasing need for high-capacity data communications is driving the development and implementation of ATM [56]. There is no doubt that data communications applications will be initially dominant in ATM networks and, moreover, will be the first test of ATM acceptance. Examples of the applications that could require ABR service include LAN interconnection, image/document retrieval, interactive data transfer, and so on. ATM networks provide connection-oriented services where data transport occurs on the virtual circuit established before transmission. However, a vast majority of existing networks and data services are inherently

connectionless which require no connection setup before transmission. In addition, the traffic generated by these applications is typically very bursty and hard to predict. The sources are likely to generate on/off traffic streams with short bursts at a speed close to maximum link rate and separated by relatively long silences. Although several studies are available describing some of their statistical characteristics, a complete description of the statistical behavior and the associated traffic models is still lacking.

The traffic is unpredictable at connection setup and thus it is unrealistic to expect the users to declare their bandwidth requirements accurately in advance of transmission. For those applications, the network is unable to provide an explicit guarantee of service since insufficient information of their traffic characteristics is known. What is needed for the data applications is a service that serves all active sources on a *best-effort* basis. The available bandwidth for this service should be dynamically allocated between the users. This kind of service is referred by the ATM Forum as the Available Bit Rate (ABR) service [56].

The burstiness of ABR traffic type covers a wide range of time scales and the peak bit rates are typically several orders of magnitude larger than the average rates. The unpredictable traffic characteristics and large burstiness make this kind of traffic ideal for statistical multiplexing where many connections cooperate to share a finite amount of bandwidth.

The traffic behavior of the various existing and foreseeable data applications depends strongly on the usage patterns of the applications and can thus hardly be characterized by means of a manageable set of parameters. Their target service requirements also rely on applications, high layer protocols and source characteristics. For instance, animation graphics may produce long traffic bursts with peak rates larger than 10 Mbps and will need QOS requirements similar to that required for real time video. Electronic mail generates individual messages with sizes from less than 1

kbyte up to several Mbytes but can tolerate delay for up to minutes. Target cell loss probabilities for these applications also cover a wide range and their impact usually depends on upper layer protocols. These protocols generally rely on retransmission to recover from information loss.

For the applications of critical data transfer such as image retrieval or very high speed data services, it is necessary to perform burst admission control. The Fast Reservation Protocol can be used to ensure that the data bursts will not be transmitted unless sufficient resources are available.

Many data applications are relatively tolerant to delay but sensitive to cell loss. If there is no flow control mechanism that allows the network to control the cell emission process at traffic sources, we will require a very large buffer at ATM switches to absorb the cells of simultaneously arriving bursts so that a reasonable cell loss rate can be achieved. It has been proposed to use closed-loop feedback control schemes for ABR service to modulate traffic streams based on current network status.

Table 4.1 summarizes the important traffic characteristics, requirements and applicable control functions for the three basic ATM service categories.

4.3 Traffic Modeling

A key element in simulating or analyzing communication networks is traffic modeling. Traffic models are used in modeling technology either as part of an analytical model or to drive a discrete-event simulation. Inappropriate traffic models could lead to poor prediction of network performance and may misdirect design and management decisions. In order to obtain a successful performance evaluation, a clear understanding of the nature of the traffic in the target system and selection of a suitable random traffic model are crucial. This section presents a review of the commonly used traffic models for different types of traffic.

Table 4.1: Traffic characteristics and applicable control functions for ATM service categories

ATM Service Categories / Characteristics & Control Functions	Constant Bit Rate (CBR)	Variable Bit Rate (VBR)	Available Bit Rate (ABR)
Applications	Voice trunk and Traditional video channels	Video conference and Distribution video	LAN interconnection, Image/document retrieval Interactive data transfer and X window applications
Connection Admission Control	Peak bandwidth allocation	Statistical bandwidth allocation	None (possibly burst admission control)
QOS Requirements	Very low cell loss and cell delay variation	Committed cell loss and delay	Not specified
In-Call Congestion Control	None	None	Yes
Buffering Requirements	Small (~ 100 cells)	Small to medium (< 1000 cells)	Large (> 1000 cells, depending on Round Trip Time and link bandwidth)
Statistical Models	Periodic arrival process, IPP (single voice source), MMPP (aggregated traffic), Poisson/Bernoulli (aggregated traffic)	AR (video traffic), Markov process (video traffic), Fluid traffic models, TES models, Poisson/Bernoulli (aggregated traffic)	Poisson/Bernoulli, Self-similar models, Packet train models

Many traffic models that are capable of capturing the autocorrelated nature of voice or video sources have been proposed and studied [31, 59]. The *Interrupted Poisson Process* (IPP) [102, 60], which is also known as on-off model and is essentially a two-state *Markov-Modulated Poisson Process* (MMPP) with zero arrival rate in one state, has been widely used to describe a single voice source. The arrival process of an IPP is characterized by two alternating states; the *on* state corresponds to a talk spurt, and the *off* state corresponds to a silence. The process stays in a state for an exponentially distributed holding time and then transits to the other state. The general MMPP is a doubly stochastic Poisson process where the rate process is controlled (modulated) by its current state and in a given state, arrivals occur according to a Poisson process with a non-negative rate. The aggregated cell arrival process of independent voice sources can be modeled by a birth-death process [31], where the aggregated arrival rate is determined by the number of voice sources in the *on* state, or by a two-state MMPP model [102], where each state is associated with a positive Poisson rate.

The statistical nature of a compressed video source displays a significant diversity from a voice source. For VBR-coded video sources, since the signal is compressed by encoding only the differences between successive frames, the frame bit rate within a video scene varies very little. Only a change of scene or background of the picture can cause the frame bit rate to vary abruptly. The cell generation process within a video scene can be modeled by an *autoregressive* (AR) process [61], or by a *discrete-state continuous-time Markov process* [62]. The AR process is commonly used to drive a simulation model while the Markov process is usually employed in queueing analysis because it is more analytically tractable than the AR model. On the other hand, the scene changes can be modeled by some modulating mechanism, such as a discrete-state continuous-time Markov process with batch arrivals [63].

Other common approaches for modeling variable rate traffic sources include *fluid traffic models* [64] and *Transform-Expand-Sample* (TES) models [65]. The fluid traffic model, instead of counting traffic units individually, views traffic as a stream of fluid characterized by a flow rate. The approach provides important advantages such as comparable accuracy and enormous savings in computing. The TES models, which can be used to generate synthetic streams of realistic traffic to drive simulations, attempts to capture both marginals and autocorrelations of empirical data simultaneously.

Poisson processes have a long history in characterizing data traffic in computer networks, because of their relative mathematical simplicity and elegant analytical properties. In addition, it is well known that in traffic applications that physically comprise a large number of independent traffic streams, the superposition process can be approximated by a Poisson process (or a Bernoulli process for discrete-time cases) [66]. In order to capture the strong autocorrelations of bursty traffic that is expected to dominate broadband networks, several new traffic models, which are based on high-quality high-resolution traffic measurements in real network environments, are currently under study. Some well-known examples of the traffic models include the *self-similar models* [67] and *packet train models* [103], and they can be employed to generate long traces of synthetic traffic for simulations. A summary of the commonly used traffic models for different ATM traffic classes is also presented in Table 4.1.

4.4 Traffic Management Mechanisms

The main objective of ATM traffic management is to achieve resource efficiency improvement while supporting the QOS requirements of users in both normal and overload conditions. As was described in Section 2, the ATM service categories have distinct service demands and require a set of control functions to prevent performance

degradation during congestion. This section discusses in further detail the roles of these traffic management mechanisms for the support of each service category.

4.4.1 Network Resource Engineering

The major function of Network Resource Engineering is to allocate the finite network resource among different ATM services so as to achieve an acceptable level of connection blocking performance. To perform this control function successfully, we will require knowledge of the connection characteristics, such as predicted arrival rates, expected connection blocking probability, call durations, usage characteristics and bandwidth requirements. The finite link bandwidth is then logically divided between service categories based on these estimates to provide satisfactory resource provision and determine acceptable loading levels. The available bandwidth capacity and utilization for a service type will be used by the connection admission control as the basis for admitting or rejecting a new connection.

For users of CBR or VBR services that require QOS agreement sustained on a per-connection basis, the size of available network bandwidth affects directly the connection blocking performance of the service type. For ABR applications where bandwidth is not reserved explicitly on a per connection basis, all connections will share the unused portion of network bandwidth. However, a minimum amount of bandwidth should be provisioned for the service category to ensure a very low probability of congestion. In addition, the division of network bandwidth capacity should be dynamically adjusted to optimize resource efficiency as traffic mix changes.

4.4.2 Connection Admission Control

Connection admission control determines if a new request will be admitted to the network and reserves network resources at the connection establishment, so that consistent performance can be received during the lifetime of the connection. Thus the central issue of the connection admission control is to decide if a new connection

can be safely multiplexed with the existing traffic loading without leading the network into an unacceptable level of congestion. The decision to accept or deny a connection request will be based primarily on the bandwidth and QOS requested by the source and the available bandwidth for a given service category.

Allocating bandwidth at connection peak rate is the simplest control approach and is particularly suitable for the applications of CBR service. A new connection is accepted only if the required peak bandwidth is available at each link along the connection's selected path. This approach offers a very strong performance guarantee and is relatively easy to implement. However, it makes poor use of the network resource when the connection's peak rate is much higher than its average rate.

Connections of VBR service can be multiplexed statistically to take advantage of the variable rate nature of individual applications and achieve some gain in resource efficiency. By considering the traffic characteristics of connections and estimating the probability of consequential congestion, the network may accept a number of connections the sum of whose peak rates exceeds the link capacity. In general, an effective bandwidth is computed and allocated based on a set of traffic descriptors specifying the variability of the bit rate, which may include peak cell rate, sustainable cell rate and burst tolerance.

For ABR service in which it might not be possible to estimate the average rate in advance, under normal conditions the connection admission control may admit the request without allocating any bandwidth. All the connections dynamically adapt themselves to current network status by sharing the minimum provisioned bandwidth for this service type and the unused bandwidth left from other service categories.

4.4.3 Explicit Congestion Notification

Explicit Congestion Notification (ECN) is a reactive feedback control mechanism that controls the rate at which each source emits cells into the network on

every virtual connection by using congestion indicators [57]. The congestion indicators are generated at the congested node and they can be sent in the forward direction (FECN) or in backward direction of the path (BCEN). In FECN when a path through a switching node becomes congested, the congestion information is transferred to the destination node by setting the explicit forward congestion indicator bit with the header of ATM cells to 1. After receiving the information, the destination end system sends congestion notification cells back to the traffic source to indicate the congestion status. The source will then dynamically adjust its cell transmission rate based on the feedback. On the other hand, in BECN congestion notification is returned directly from the congested node back to the source. Obviously the BECN control scheme is able to react to network congestion faster than the FECN, although it requires more hardware in the switch for implementation.

It has been shown that ECN is an effective control scheme to reduce information loss and improve network throughput during congestion periods for some data applications such as LAN interconnection.

4.4.4 Credit-based Flow Control

The control objective of credit-based flow control is similar to the ECN scheme described in the previous section. However, instead of controlling the source's cell transmission rate based on congestion indicators from the congested node, credit-based flow control manipulates the traffic being sent to the receiving end of a link according to the available resources for the virtual connection (i.e., credits). The credit approach is thus a link-by-link per-virtual-connection flow control mechanism where each link performs the scheme independently. The sending end of a link is allowed to transmit cells to its receiving end only if it has credits for the virtual connection. A credit balance is maintained at the sending end and is updated periodically by credit cells sent by its receiver [68].

The credit-based approach has the advantage of excellent bandwidth utilization with zero cell loss in a high-loaded network. When multiple connections are active, the available bandwidth is shared fairly between them. If a connection cannot fully utilize its share, the unused bandwidth will be acquired quickly by other aggressive connections. However, the control scheme requires per-virtual-connection buffering and processing in every switch and thus increases implementation complexity on the switch hardware.

4.4.5 Burst Admission Control

The Burst Admission Control is a control procedure that allows users to negotiate bandwidth with the network during a connection. This approach should be applied to the applications that need spontaneous delivery of long bursts of critical data so that satisfactory QOS can be guaranteed during data transfer. With this control scheme, the bandwidth will be allocated only to the bursts based on the peak bandwidth requirements prior to their transmissions. If sufficient bandwidth cannot be reserved on some link of the connection path, the transmission of the burst will be blocked or delayed. In order to ensure an acceptable probability of burst blocking, the network may impose a limit on the connection peak rate that can be supported by a link, for instance, 10 percent of the link bit rate.

Blocking large data transfers at the source during network overload can prevent performance degradation to other connections and thus improve network stability under high loading.

4.4.6 Usage Parameter Control

For each connection that has been admitted to the network, there should be a service contract specifying a set of service parameters, including bandwidth parameters and QOS requirements. With the service agreement, a set of mechanisms for monitoring traffic compliance during the transport of ATM cells is required. This

control function provides stable operation of the network and resource protection for compliant users by discarding or tagging for priority discard the traffic that violates its specified parameters.

The traffic control mechanism is performed by monitoring continuously a single indicator, the cell loss priority (CLP), in every ATM cell header of a virtual connection. If the CLP bit of a cell is set to 1, it indicates that the cell carries nonessential information and may be discarded preferentially when congestion is encountered along the path. Another function served by the indicator is that the network provider can mark the excessive cells that are not in compliance with the traffic limits by setting their CLP indicators to 1. Thus those cells will be transferred at their own risk to ensure fairness between competing users.

Generally, the peak cell rate of a connection should be monitored for all kinds of applications. Additional functionality is required for those applications in which the network bandwidth is not reserved at their peak rate. The traffic monitoring mechanism established for these connections should effectively impose an upper bound on the worst case traffic behavior expected from the connection.

4.4.7 Priority Control

The priority control which determines how cells should be treated in the network is an important control function to support different service demands of ATM users. During connection establishment, each connection should be assigned with explicit priorities according to the service category and service requirements. The priority levels of a connection are registered in the virtual connection identifier (VCI) table so that the priorities of an incoming cell can be identified.

Two different types of the connection-level priority should be specified: the delay priority and the loss priority. A traffic scheduling policy, which determines the serving sequence between different delay priorities and how a finite outgoing link bandwidth should be allocated in switching nodes, provides effective controls on

network queueing delay and delay variation for a given cell. On the other hand, the loss priority of a connection is used by a cell loss mechanism that manages the accesses to finite buffers in a switching node to satisfy various connection loss requirements.

In addition to the overall traffic control for different connection-level priorities, additional thresholds imposed on the buffer utilization are required to support different cell loss rates of the cell-level priority (i.e. cell loss probabilities for the cells of the same connection with CLP=0 and CLP=1). The cells with lower priority will be discarded when the buffer utilization exceeds the threshold to provide better cell loss performance for the cells carrying essential information.

Figure 4.2 shows an architecture of the ATM traffic control. Finally, Table 4.2 summarizes the performance objectives and control functions of the important traffic management schemes that were discussed in this chapter.

In the following chapters, three (3) traffic management schemes are investigated and optimized with respect to critical network resources. A *space priority buffer management* scheme, which controls the finite buffer in a switching node to support an arbitrary number of priority classes, is analyzed in Chapter 5. A multinomial traffic model is chosen to give a good approximation of the aggregated input traffic since each priority class may consist of a large number of independent traffic sources. Previous works have typically considered only two priority classes and overly simplified queueing models. In Chapter 5 we present a general approach for multiple priority classes and optimization procedures for the optimal choice of loss thresholds. In Chapter 6, two feedback flow control schemes for ABR traffic, the BECN and the credit-based mechanisms, are investigated. The worst-case (e.g., long file transfers with TCP/IP) network performance with and without the flow controls is examined. We show that with the implementation of the flow control schemes, packet retransmission process, which might degrade severely the network throughput, can be greatly reduced or completely eliminated, thus providing a substantial improvement

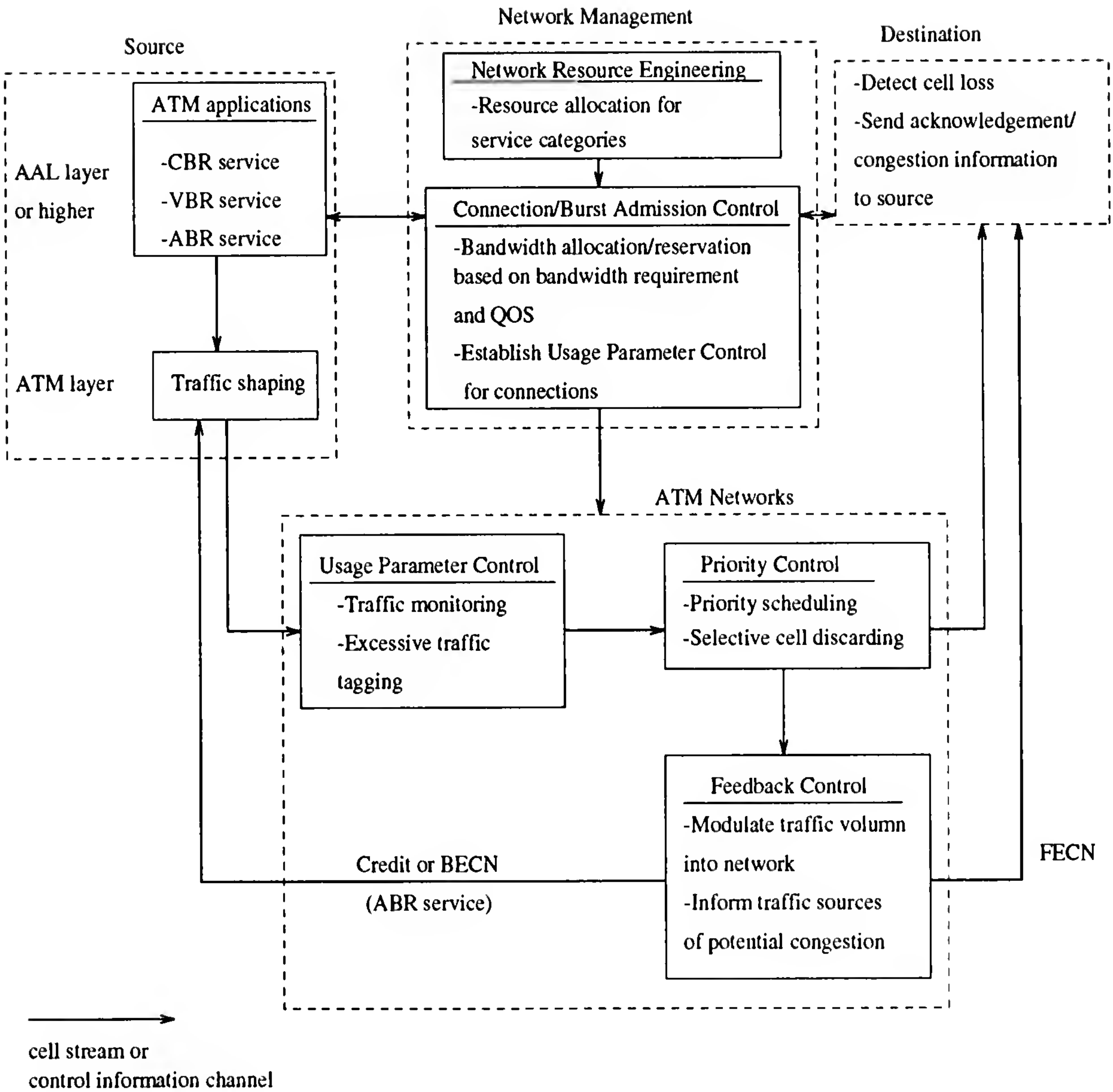


Figure 4.2: ATM Traffic Control Architecture

Table 4.2: ATM Traffic Management

Traffic management	**ATM service type**	**Control time scale**	**Congestion control**	**Performance objective**	**Control functions**
Network resource engineering	CBR, VBR, ABR	long term or whenever traffic condition changes	preventive	- resource management for acceptable call blocking performance	- resource allocation for different service categories
Connection admission control	CBR, VBR, ABR	call duration	preventive and reactive	- CBR and VBR: acceptable connection blocking probability - ABR: no committed connection blocking probability	- CBR: peak bandwidth allocation for connection - VBR: statistical bandwidth allocation for connection - ABR: no bandwidth allocation at connection setup
Explicit congestion notification (forward or backward)	ABR	burst duration or round trip delay time	reactive	- committed cell loss and delay	- rate-based flow control - inform traffic sources of potential congestion
Credit-based flow control	ABR	burst duration or round trip delay time	preventive	- no cell loss	- link-by-link flow control - per connection buffering and traffic management
Burst admission control	ABR	burst duration	preventive	- no committed burst blocking probability - committed cell loss and delay for accepted burst	- peak bandwidth allocation for burst
Usage parameter control	CBR, VBR, ABR	cell time	preventive	- CBR: very low cell loss and cell delay variation - VBR: committed cell loss and delay - ABR: acceptable cell loss and delay	- peak rate enforcement - excessive traffic tagging
Priority control	CBR, VBR, ABR	cell time	preventive and reactive		- priority assignment for connections - selective cell discarding under congestion

on the network performance. In addition, a slow-start technique that can improve the performance of the rate-based control scheme to a comparable level as with the credit-based approach is proposed. We conduct a complete analysis on the issues of performance improvement, resource efficiency, and practical implementation of these control mechanisms.

CHAPTER 5
SPACE PRIORITY BUFFER MANAGEMENT FOR ATM OVERLOAD CONTROL

5.1 Buffer Management Schemes for ATM Switches

In this chapter, the issue of effectively managing the finite output buffers of an ATM switch is concerned. Due to the bursty and unpredictable nature of some traffic sources, the occurrence of buffer overflows is inevitable. Simply dimensioning the resource to satisfy the most stringent service requirement is generally not a practical solution and will lead to low efficiency. In addition, systems with a large buffer can introduce significant cell delay variation for the passing traffic if no priority control is applied. This situation may be unbearable for the delay-sensitive applications. More dynamic controls such as prioritizing different traffic according to their service requirements are necessary for ATM systems to achieve high resource utilization and ensure adequate network performance.

There are two major categories of the priority queueing mechanisms for finite buffer systems being proposed: the *time priority* and the *space priority* strategies. The time priority strategies provide preferential treatment to the traffic with critical delay requirements while the space priority favors the traffic with sensitive loss requirements. Although these two kinds of queueing strategies were commonly studied and analyzed in an independent manner, they can be implemented cooperatively in a finite buffer system without conflict. For instance, a space priority (buffer access) control scheme can be implemented at the server for controlling buffer memory input to ensure satisfactory loss rates for different loss priority classes, while a time priority (scheduling) control scheme can be enforced at the server for controlling buffer memory output to handle various delay requirements. Thus the delay and loss priorities

of a virtual connection can be assigned in a more flexible way to accommodate the diverse QOS of ATM applications.

Two space priority queueing strategies, *Push Out* scheme and *Partial Buffer Sharing* (PBS) scheme, have been suggested. In the push out scheme, an arriving high priority cell is allowed to take the place of any low priority cell in the queue when the buffer is full. If there is no low priority cell in the queue, the arriving high priority cell is discarded. On the other hand, in the partial buffer sharing scheme, a threshold is imposed on the buffer space available to both classes. A low priority cell is admitted to the system only if the current queue length is less than the threshold, otherwise it is lost. Obviously the push out scheme provides better resource efficiency, since a newly arriving cell is rejected only if the buffer is saturated. However, in order to keep track of cell positions and preserve proper cell sequencing, the push out scheme implementation requires a much more complicated buffer management logic, which may be undesirable for systems with large buffers. For instance, the egress buffer memory in an egress adaptor of the CPS-100 switch described in Chapter 3 can buffer up to 125,000 cells. The complexity to implement the push out scheme in such a large buffer could be unacceptable; thus the partial buffer sharing scheme may be a better choice if a large buffer space is desirable and the buffer efficiency is not a critical consideration.

Recently much attention has been spent on the analysis and performance evaluation of the partial buffer sharing mechanism [69, 70, 72, 73, 74]. It has been shown that the effectiveness of the priority discarding scheme can be very substantial, as compared with the system without priority control. Most of the previous studies consider only two priorities, in which a single threshold is set on the buffer utilization to determine if an arriving cell with low cell-level priority (i.e., CLP bit set to 1) should be discarded. One exception concerning an arbitrary number of priorities can be found in [69]. In that study the problem of selecting an optimal set of nested

thresholds was considered based on a simplified queueing model. In order to enforce more dynamic control and to accommodate the user's varied loss requirements, it is desirable that more loss priority classes can be supported by a shared buffer system. For instance, at a switching node a shared buffer memory may be allocated to accommodate CBR and VBR traffic. The network provider may specify several grades of service that offer different cell loss rates for a number of connection-level priorities to support various CBR and VBR applications. In such a system, we need to determine multiple thresholds on the buffer so that the cell loss requirements can be satisfied.

To prevent excessive cell delay occurring to real-time traffic at a switching point with a large buffering space, at least two logical buffers should be allocated to separate real-time traffic such as CBR and VBR from non-real-time traffic such as ABR. The real-time traffic will require only a small buffer since the queue will always be served in a higher priority. The PBS mechanism can be applied to the real-time buffer to provide a strict control on cell loss requirements and queueing delay introduced during cell buffering. The major portion of the buffering space could be assigned to the non-real-time traffic for trading off cell loss for delay. The congestion problem for this class of traffic can be controlled effectively by the feedback flow control schemes, which will be investigated in the next chapter.

In this chapter an accurate queueing model to characterize the system is developed and methods to optimize the system performance are presented. This research is motivated by the fact that in a heterogeneous environment, loss requirements and traffic mixes can be very diverse and it is important to dimension the finite buffer space properly to guarantee acceptable loss probabilities for different classes. Moreover, simulating real buffer systems on computers to evaluate whether the loss probabilities satisfy given loss constraints is generally computational feasible only when the loss constraints are high (for instance $\geq 10^{-4}$). Here an analytic method to estimate loss probabilities of multiple loss classes is developed and the numerical

optimization procedures to select the best threshold levels efficiently under different system conditions are introduced.

The rest of this chapter is organized as follows. In Section 5.2, a complete description of the problem is presented. A detailed queueing model analysis, along with the numerical results, is addressed in Section 5.3. The system optimization procedures to search for the best threshold levels under different system conditions are introduced in Section 5.4. Finally, a conclusion is given in Section 5.5.

5.2 Partial Buffer Sharing Scheme and Problem Statement

The research is concerned with a space priority queueing strategy for the finite output buffer in ATM switching systems. The buffer implemented in the output of a switch is used to temporarily store those cells which cannot be immediately sent out. The instantaneous buffer access rate is designed to be much higher than the buffer output rate to accommodate simultaneous arrivals from different inputs. To support the diverse loss QOS requirements of ATM services, the buffer access is controlled by a generalized PBS space priority queueing strategy. In the PBS a set of loss thresholds is imposed on the buffer space available to different priority classes. Each loss threshold B_i ($i = 1, ...N$) is associated with a loss priority class i. A class i arrival can be admitted to the system only if the current buffer state is less than its designated threshold, otherwise the arriving cell will be discarded. If we assume that priority N represents the highest priority class and priority 1 is the lowest priority class, priority N cells will have access to the whole buffer and, therefore, $B_N = B$ and $0 < B_1 < B_2 < \cdots < B_N$. Consequently, a finite buffer can be divided into N portions: the first B_1 buffer positions are accessible for all incoming cells, the next $B_2 - B_1$ positions are accessible for class 2 or higher priority cells, and so on. The term *eligible group* will be used to denote those classes which are eligible for a specific portion of buffer space, for example, the eligible group of the first B_1 buffer positions

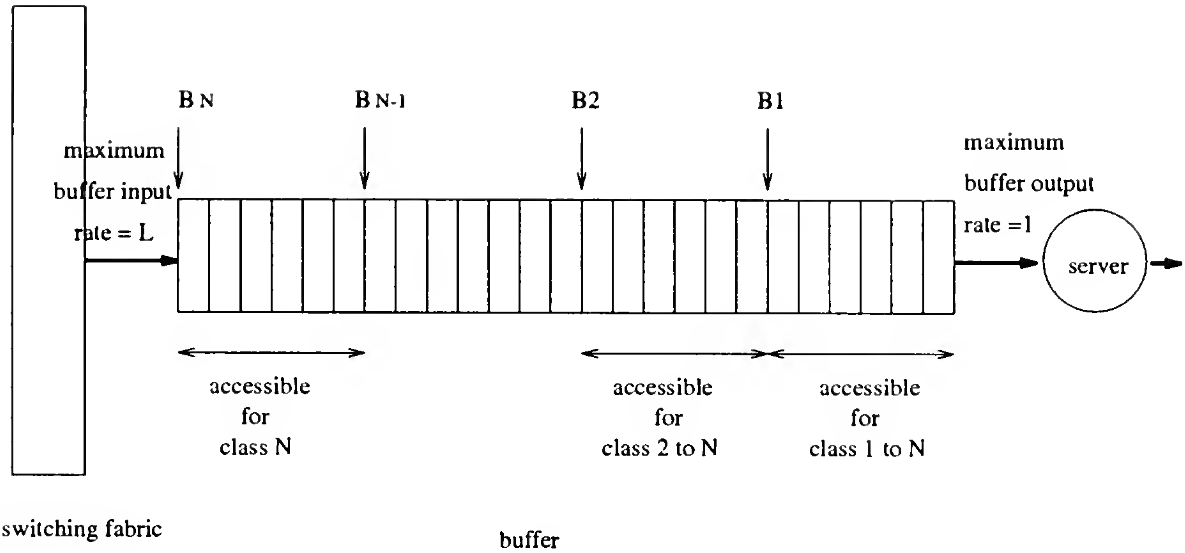


Figure 5.1: Partial Buffer Sharing Mechanism

includes all classes (eligible group 1), the eligible group of the next $B_2 - B_1$ positions consists of class 2 to class N (eligible group 2), etc. Figure 5.1 describes the PBS mechanism. It is assumed that the service discipline of the buffer is first-come-first-serve (FCFS). By adjusting the threshold values, different loss requirements under various traffic load conditions can be fulfilled.

This approach leads to the problem of how to choose a set of optimal loss thresholds under given constraints of load conditions and loss probabilities. Under a fixed load condition, an increase in the loss threshold of a particular class can decrease the loss probability of the class and also increase the loss probabilities of higher priority classes. Moreover, as stated in [69], the constraints of loss probabilities implicitly place an upper bound on the maximum load that can be supported by the system. The problem of interest is to determine a set of optimal loss thresholds such that the admissible load (ρ) to the system is maximized subject to the given traffic conditions and loss constraints. That is, the optimization problem can be formulated

as follows:

$$\text{Maximize} \quad \rho(B_1, B_2, \cdots, B_N) \tag{5.1}$$

$$\text{subject to} \quad \left\{ \begin{array}{ll} B_i, & \\ PL_i & = \text{loss constraint of class } i, \\ r_i & = \text{ratio of class } i \text{ traffic load to total traffic load.} \end{array} \right\}$$

It is seen that the set of optimal loss thresholds under a maximum admissible load might not be unique, that is, there may exist more than one set of loss thresholds which give different loss probabilities but which may still fulfill the loss constraints. In those cases, the optimal thresholds will be taken to be those B_i which satisfy (5.1) and the following condition: for each loss priority class, the B_i is the smallest number that satisfies the loss constraint of the class. Thus the buffer space retained for class N will be the largest and the loss probability of class N will be the lowest among all possibilities.

To solve the problem we need to develop a queueing model to characterize the system.

5.3 A Queueing Model of Partial Buffer Sharing Mechanism

Based on the slotted nature of fixed-length cell processing in ATM switches, the queueing system can be modeled as a discrete time Markov chain with finite buffer capacity B and N classes of arrivals. The Partial Buffer Sharing (PBS) scheme controls access to the shared buffer according to the loss priority status of incoming cells. A queueing model similar to the one stated by Lin and Silvester [70] is developed. The queueing model presented in their work predicts accurately the loss probabilities of a binary system with a given loss threshold and a multinomial traffic distribution. However, unlike their model, the queueing model constructed here is intended to describe a multiple-threshold (N-class) system instead of a single-threshold (two-class) system.

Suppose the transmission time of an ATM cell is one timeslot and all arrivals and departures occur at slot boundaries. It is assumed that incoming cells have to wait at least one timeslot before being transmitted. Moreover, assume that cells arrive at the system in a batch. The number of arrivals in a batch is random and upper-bounded by L. The parameter L can represent the number of input ports or the maximum number of cells that can be written into buffer within a timeslot. The numbers of class i ($i = 1, ...N$) cells in the batch are described by a joint probability mass function (pmf), $a_{1,2,...N}(n_1, n_2, ...n_N)$ = probability of n_1 class 1 cells, n_2 class 2 cells, etc., in an arriving batch. It is assumed that the probability of a class i cell appearing in a batch is p_i (i.e., $p_i = \rho r_i$) and is statistically independent of past arrivals. Then the joint pmf $a_{1,2,...N}(n_1, n_2, ...n_N)$ can be written as a multinomial pmf:

$$\begin{aligned} a_{1,2,...N}(n_1, n_2, ...n_N) &= \binom{L}{n_1, n_2, ..., n_N, (L - n_1 - n_2 - ... - n_N)} \\ &\quad \cdot \left(\frac{p_1}{L}\right)^{n_1} \left(\frac{p_2}{L}\right)^{n_2} \cdots \left(\frac{p_N}{L}\right)^{n_N} \\ &\quad \cdot \left(1 - \frac{p_1}{L} - \frac{p_2}{L} - ... - \frac{p_N}{L}\right)^{L-n_1-n_2-...-n_N} \end{aligned} \tag{5.2}$$

Since only class 2 or higher priority cells (i.e., eligible group 2) can be accepted as buffer state greater than B_1, we need to calculate the marginal pmf

$$\begin{aligned} a_{2,...N}(n_2, ...n_N) &= \sum_{n_1=0}^{L} a_{1,2,...N}(n_1, n_2, ...n_N) \\ &= \binom{L}{n_2, ..., n_N, (L - n_2 - ... - n_N)} \\ &\quad \cdot \left(\frac{p_2}{L}\right)^{n_2} \cdots \left(\frac{p_N}{L}\right)^{n_N} \left(1 - \frac{p_2}{L} - ... - \frac{p_N}{L}\right)^{L-n_2-...-n_N} \end{aligned} \tag{5.3}$$

and for the similar reason, the following marginal pmfs are also required:

$$\begin{aligned} a_{3,...N}(n_3, ...n_N) &= \sum_{n_1=0}^{L} \sum_{n_2=0}^{L} a_{1,2,...N}(n_1, n_2, ...n_N) \\ &= \binom{L}{n_3, ..., n_N, (L - n_3 - ... - n_N)} \end{aligned}$$

$$\cdot\left(\frac{p_3}{L}\right)^{n_3}\cdots\left(\frac{p_N}{L}\right)^{n_N}\left(1-\frac{p_3}{L}-\ldots-\frac{p_N}{L}\right)^{L-n_3-\ldots-n_N}$$

$$\vdots$$

$$a_{N-1,N}(n_{N-1},n_N) = \binom{L}{n_{N-1},n_N,(L-n_{N-1}-n_N)} \cdot\left(\frac{p_{N-1}}{L}\right)^{n_{N-1}}\left(\frac{p_N}{L}\right)^{n_N}\left(1-\frac{p_{N-1}}{L}-\frac{p_N}{L}\right)^{L-n_{N-1}-n_N}$$

The pmfs of the aggregated batch size for different eligible groups are obtained by:

$$\begin{aligned} t_1(y) &= \text{pmf of the aggregated batch size for eligible group 1} \\ &= \sum_{\text{all combinations subject to } \{0\le n_1,n_2,\ldots n_N\le y,\, n_1+n_2+\ldots+n_N=y\}} a_{1,2,\ldots N}(n_1,n_2,\ldots n_N) \\ t_2(y) &= \text{pmf of the aggregated batch size for eligible group 2} \qquad (5.4) \\ &= \sum_{\text{all combinations subject to } \{0\le n_2,\ldots n_N\le y,\, n_2+\ldots+n_N=y\}} a_{2,\ldots N}(n_2,\ldots n_N) \\ &\vdots \\ t_N(y) &= \text{pmf of the aggregated batch size for eligible group N} = a_N(y) \end{aligned}$$

To characterize the system, we need to find the steady state buffer size pmf vector $q = [q(0), q(1), \ldots, q(B)]$ by solving the stationary equation:

$$q = q\mathbf{T} \qquad (5.5)$$

where $\mathbf{T}$ is the state transition matrix. The element e_{mn} in matrix $\mathbf{T}$ specifies the state transition probability of the system going from state m to state n. The equation below shows an example of the state transition matrix $\mathbf{T}$.

$$
\left[
\begin{array}{cccccccccc}
t_1(0) & t_1(1) & t_1(2) & \cdots & t_1(L) & 0 & 0 & \cdots & \cdots & \cdots \\
t_1(0) & t_1(1) & t_1(2) & \cdots & t_1(L) & 0 & 0 & \cdots & \cdots & \cdots \\
0 & t_1(0) & t_1(1) & t_1(2) & \cdots & t_1(L) & 0 & \cdots & \cdots & \cdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & t_1(0) & t_1(1) & t_1(2) & \cdots & t_1(B_1-1) & t_{12}(B_1,0) & t_{12}(B_1,1) & \cdots \\
0 & \cdots & 0 & t_1(0) & t_1(1) & \cdots & t_1(B_1-2) & t_{12}(B_1-1,0) & t_{12}(B_1-1,1) & \cdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \cdots \\
0 & 0 & \cdots & & & & t_1(0) & t_{12}(1,0) & t_{12}(1,1) & \cdots \\
0 & 0 & \cdots & & & & 0 & t_2(0) & t_2(1) & \cdots \\
0 & 0 & \cdots & & & & 0 & 0 & t_2(0) & t_2(1) \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots \\
0 & 0 & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots
\end{array}
\right.
$$

$$
\left.
\begin{array}{cccccc}
\cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
t_{12}(B_1, B_2-B_1-1) & t_{123}(B_1, B_2-B_1, 0) & \cdots & \cdots & \cdots & 0 \\
t_{12}(B_1-1, B_2-B_1-1) & t_{123}(B_1-1, B_2-B_1, 0) & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
t_{12}(1, B_2-B_1-1) & t_{123}(1, B_2-B_1, 0) & t_{123}(1, B_2-B_1, 1) & \cdots & t_{1234}(1, B_2-B_1, B_3-B_2, 0) & \cdots \\
t_2(B_2-B_1-1) & t_{23}(B_2-B_1, 0) & t_{23}(B_2-B_1, 1) & \cdots & \cdots & \cdots \\
t_2(B_2-B_1-2) & t_{23}(B_2-B_1-1, 0) & t_{23}(B_2-B_1-1, 1) & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & t_N(0) & t_N(1)
\end{array}
\right]
$$

Since the probability of a class i cell appearing in a batch is p_i, the total traffic load to the system is $p_1 + p_2 + \cdots + p_N$ and as the buffer state exceeds the first threshold, the effective traffic load becomes $p_2 + p_3 + \cdots + p_N$. Similarly, the effective traffic load declines by a class arrival rate when the buffer state goes beyond the corresponding threshold. Moreover, the transitions from state m to state $n < m - 1$ are impossible because at most one cell will be served at each timeslot and the state transition probabilities from state m to state $n > m + L - 1$ ($n > L$ for $m = 0$) are also zeros since the number of arrivals in a batch is upper-bounded by L. The denotations of other state transition probabilities are explained as follows. The $t_i(y_i)$ represents the probability that y_i cells from eligible group i in an arriving batch enter the system, the $t_{ij}(y_i, y_j)$ is the probability that the first y_i cells from eligible group i and exactly y_j cells thereafter from eligible group j in an arriving batch enter the system, and so on. The transitions between state $0 < m \le B_i \mid_{i=1}$ and $0 \le n < B_i \mid_{i=1}$ or state

$B_{i-1} < m \leq B_i$ and $B_{i-1} \leq n < B_i$ depend only on the arrivals of eligible group i and their probabilities are equal to $t_i(n+1-m)$.

For the transitions from state $0 < m \leq B_i \mid_{i=1}$ to $B_i \mid_{i=1} \leq n < B_{i+1}$ or state $B_{i-1} < m \leq B_i$ to $B_i \leq n < B_{i+1}$, the probabilities depend on the arrivals of eligible group i and $(i+1)$ as well as the cell positions in the batch. Assuming every permutation of the cell positions is with equal likelihood, the probabilities $t_{i,(i+1)}(y_i, y_{(i+1)})$ can be computed by considering the conditional probability that, given a number of cells from eligible group i, there are exactly $y_{(i+1)}$ cells from eligible group $(i+1)$ between position $y_i + 1$ and the last position in the arriving batch [70]. In the next step we multiply the conditional probability with the marginal pmf $a_{i,\ldots N}(n_i, \ldots n_N)$. The $t_{i,(i+1)}(y_i, y_{(i+1)})$ can be obtained by summing up all these probabilities of possible combinations of $n_i, n_{i+1}, \ldots n_N$. For example, the corresponding probabilities for a four-class system ($N = 4$) are given by:

$$t_{12}(y_1, y_2) = \sum_{\text{all combinations subject to } \{n_2+n_3+n_4 = y_2\}}^{y_1+y_2} \Big\{ \sum_{n_1 = y_1+y_2-(n_2+n_3+n_4)}^{L-(n_2+n_3+n_4)} \cdot a_{1,2,3,4}(n_1, n_2, n_3, n_4) \frac{\binom{n_1+n_2+n_3+n_4-y_1}{y_2}\binom{y_1}{n_2+n_3+n_4-y_2}}{\binom{n_1+n_2+n_3+n_4}{n_1}} \Big\} \quad (5.7)$$

$$t_{23}(y_2, y_3) = \sum_{\text{all combinations subject to } \{n_3+n_4 = y_3\}}^{y_2+y_3} \Big\{ \sum_{n_2 = y_2+y_3-(n_3+n_4)}^{L-(n_3+n_4)} a_{2,3,4}(n_2, n_3, n_4) \cdot \frac{\binom{n_2+n_3+n_4-y_2}{y_3}\binom{y_2}{n_3+n_4-y_3}}{\binom{n_2+n_3+n_4}{n_2}} \Big\} \quad (5.8)$$

$$t_{34}(y_3, y_4) = \sum_{n_4=y_3}^{y_3+y_4} \Big\{ \sum_{n_3=y_3+y_4-n_4}^{L-n_4} a_{3,4}(n_3, n_4) \cdot \frac{\binom{n_3+n_4-y_3}{y_4}\binom{y_3}{n_4-y_4}}{\binom{n_3+n_4}{n_3}} \Big\} \quad (5.9)$$

Depending on the distance between thresholds, the state transition probabilities from $0 < m \leq B_1$ to $n > B_2$ or state $B_{i-1} < m \leq B_i$ to $n > B_{i+1}$ might need

to be solved. For instance, we need to figure out the probability of $t_{123}(y_1, y_2, y_3)$ given the first y_1 cells from eligible group 1 and the following $y_2 (y_2 = B_2 - B_1)$ cells from eligible group 2 and y_3 cells afterwards from eligible group 3 in an arriving batch if $y_1 + y_2 + y_3 \leq L$. Unfortunately, it is generally too difficult to compute these probabilities analytically and thus they are solved in a numerical way. The means to compute those transition probabilities across more than one threshold are similar to the way that has been done for computing $t_{i,(i+1)}(y_i, y_{(i+1)})$, except that the conditional probability is now evaluated numerically. Finally, the probabilities in any row of the state transition matrix need to be normalized so that the sum of all elements in a row is equal to 1.

Once these state transition probabilities are determined, the equation $q = q\mathbf{T}$ can be solved recursively to obtain the equilibrium state probabilities $q(k)$. Specifically, this can be done by assuming an initial value for $q(0)$ to solve for $q(1), q(2), ..., q(B)$ sequentially and later deducing $q(0)$ from the condition $\sum_{k=0}^{B} q(k) = 1$. This method is chosen because of easy programming. The stationary equation can also be solved by many other numerical solution techniques.

5.3.1 Calculation of Loss Probabilities

The state probabilities $q(k)$ can be used to calculate steady state loss probabilities for different priority classes. To formulate the loss probabilities, the results of a single-threshold system derived by Lin and Silvester [70] is adapted to a multiple-threshold system. A more detailed explanation for the derivation of the loss probabilities can be found in their paper.

Considering the probability for a tagged class i cell to be able to enter the system, since it is assumed that the number of arrivals in a batch is upper-bounded by L, the probability is always equal to 1 if the current buffer state k is less than or equal to $B_i - L + 1$ (if $B_i - L + 1 \geq 0$). On the other hand, if the buffer state k is between B_i and $B_i - L + 1$ (i.e., $B_i - L + 1 < k \leq B_i$), the probability for a tagged

class i cell to successfully enter the system is related to its position as well as the positions of other class cells in the batch. To begin, let us compute the steady-state probability that a tagged class 1 cell in an arriving batch will be lost. Assuming that the occurrence of a tagged cell in each position of an batch are equally possible, the loss probability of a tagged class 1 cell (ϕ_1) can be given by

$$\begin{aligned}
\phi_1 &= 1-\overline{\phi_1}(\text{the probability that the tagged class 1 cell enter the system}),\\
\overline{\phi_1} &= \sum_{k=0}^{B_1-L+1} q(k)\{ifB_1-L+1\geq 0\}\\
&+ \sum_{k=max\{0,B_1-L+2\}}^{B_1} q(k)\\
&\cdot P[\text{the tagged class 1 cell arrives in one of the first } (B_1-k+1) \text{ positions of a batch } \mid q=k]\\
&= \sum_{k=0}^{B_1-L+1} q(k)\{ifB_1-L+1\geq 0\}\\
&+ \sum_{k=max\{0,B_1-L+2\}}^{B_1} q(k)\\
&\cdot[\sum_{\text{all combinations subject to } \{1\leq n_1+n_2...+n_N\leq L, n_1\geq 1\}} \frac{n_1\cdot a_{1,2,...N}(n_1,n_2,...n_N)}{p_1}\\
&\cdot\Omega_1(n_1,n_2,...n_N,k)]
\end{aligned} \tag{5.10}$$

where

$$\Omega_1(n_1,n_2,...n_N,k)|_{k\neq 0} = \begin{cases} 1 & \text{if } n_1+n_2+...+n_N\leq B_1-k+1 \\ \frac{B_1-k+1}{n_1+n_2+...+n_N} & \text{otherwise,}\end{cases} \quad \text{and}$$

$$\Omega_1(n_1,n_2,...n_N,0) = \begin{cases} 1 & \text{if } n_1+n_2+...+n_N\leq B_1 \\ \frac{B_1}{n_1+n_2+...+n_N} & \text{otherwise.}\end{cases}$$

The analysis of the loss probability of a class 2 cell is more complicated than the computation of class 1 loss probability because of threshold B_1 limiting the entrance of class 1 cells. Depending on the buffer state, only cells in the eligible group can be accepted to the system and thus the loss probabilities with different buffer states need to be considered separately. Using the same assumption of an uniformly

distributed position, the loss probability of a tagged class 2 cell (ϕ_2) can be formulated by

$$
\begin{aligned}
\phi_2 &= 1 - \overline{\phi_2}(\text{the probability that the tagged class 2 cell enter the system}),\\
\overline{\phi_2} &= \sum_{k=0}^{B_2-L+1} q(k)\{if B_2 - L + 1 \geq 0\}\\
&+\overline{\phi_{2,\alpha}}(\text{the probability that the tagged class 2 cell enter the system as } max\{B_2 - L + 2, B_1 + 1\} \leq k \leq B_2)\\
&+\overline{\phi_{2,\beta}}(\text{the probability that the tagged class 2 cell enter the system if } max\{0, B_2 - L + 2\} \leq k \leq B_1),
\end{aligned}
$$

$$
\begin{aligned}
\overline{\phi_{2,\alpha}} &= \sum_{k=max\{B_2-L+2,B_1+1\}}^{B_2} q(k)\\
&\cdot P[\text{the tagged class 2 cell arrives in one of the first } (B_2 - k + 1) \text{ positions}\\
&\text{of an eligible group 2 batch } |q = k]\\
&= \sum_{k=max\{B_2-L+2,B_1+1\}}^{B_2} q(k)\\
&\cdot[\sum_{\text{all combinations subject to } \{1 \leq n_2 + n_3 ... + n_N \leq L, n_2 \geq 1\}} \frac{n_2 \cdot a_{2,3,...N}(n_2, n_3, ...n_N)}{p_2}\\
&\cdot \Omega_{2,\alpha}(n_2, n_3, ...n_N, k)]
\end{aligned}
\tag{5.11}
$$

where

$$
\Omega_{2,\alpha}(n_2, n_3, ...n_N, k) = \begin{cases} 1 & \text{if } n_2 + n_3 + ... + n_N \leq B_2 - k + 1\\ \frac{B_2-k+1}{n_2+n_3+...+n_N} & \text{otherwise,} \end{cases}
$$

and

$$
\begin{aligned}
\overline{\phi_{2,\beta}} &= \sum_{k=max\{0,B_2-L+2\}}^{B_1} q(k)\\
&\cdot[\sum_{\text{all combinations subject to } \{1 \leq n_1 + n_2 ... + n_N \leq L, n_2 \geq 1\}} \frac{n_2 \cdot a_{1,2,...N}(n_1, n_2, ...n_N)}{p_2}\\
&\cdot \Omega_{2,\beta}(n_1, n_2, ...n_N, k)]
\end{aligned}
\tag{5.12}
$$

where

$$\Omega_{2,\beta}(n_1, n_2, ...n_N, k) = \begin{cases} 1 & \text{if } n_1 + n_2 + ... + n_N \leq B_2 - k + 1 \\ \frac{B_2-k+1}{n_1+n_2+...+n_N} + \sum_{n^*=B_2-k+2}^{n_1+n_2+...+n_N} & H(n^*, n_1, ...n_N, k) \text{ otherwise,} \end{cases}$$ and

$$H(n^*, n_1, ...n_N, k) = \frac{1}{n_1 + n_2 + ... + n_N}$$

$$\cdot[\sum_{r=0}^{min\{n_1-1, B_2-B_1-1\}} \frac{\binom{n^* - 1 - (B_1 - k + 1)}{r}\binom{n_1 + ... + n_N - 1 - [n^* - 1 - (B_1 - k + 1)]}{n_2 + ... + n_N - 1 - r}}{\binom{n_1 + n_2 + ... + n_N - 1}{n_2 + ... + n_N - 1}}]$$

The loss probability of a tagged class i ($i > 2$) cell can be derived in a way nearly like the loss probability of a class 2 cell. The only difference is the function H in the calculation of $\overline{\phi_{i,\beta}}$ needs to be evaluated numerically in case of $max\{0, B_i - L + 2\} \leq k \leq B_{i-2}$.

5.3.2 Numerical Results

In the previous section a queueing model is developed to compute the loss probabilities of multiple priority classes under certain threshold levels. Using the queueing model, some numerical examples showing the effectiveness of the Partial Buffer Sharing scheme are presented. Let us consider the loss probabilities of a four-class system ($N = 4$) and assume that the maximal arrivals in a batch is 10 ($L = 10$). Systems of greater priority class or larger bounds could be evaluated in a similar manner.

Impact of loss thresholds. First the effect of varying the loss thresholds is studied for a system of buffer capacity $B = 60$ and traffic load $\rho = 0.9$ with the distance between threshold levels fixed at $\triangle B = 2$. Figure 5.2 shows the loss probabilities of four priority classes as a function of threshold levels. Three different traffic mixes are considered: Figure 5.2(a) illustrates the loss probabilities of a balanced

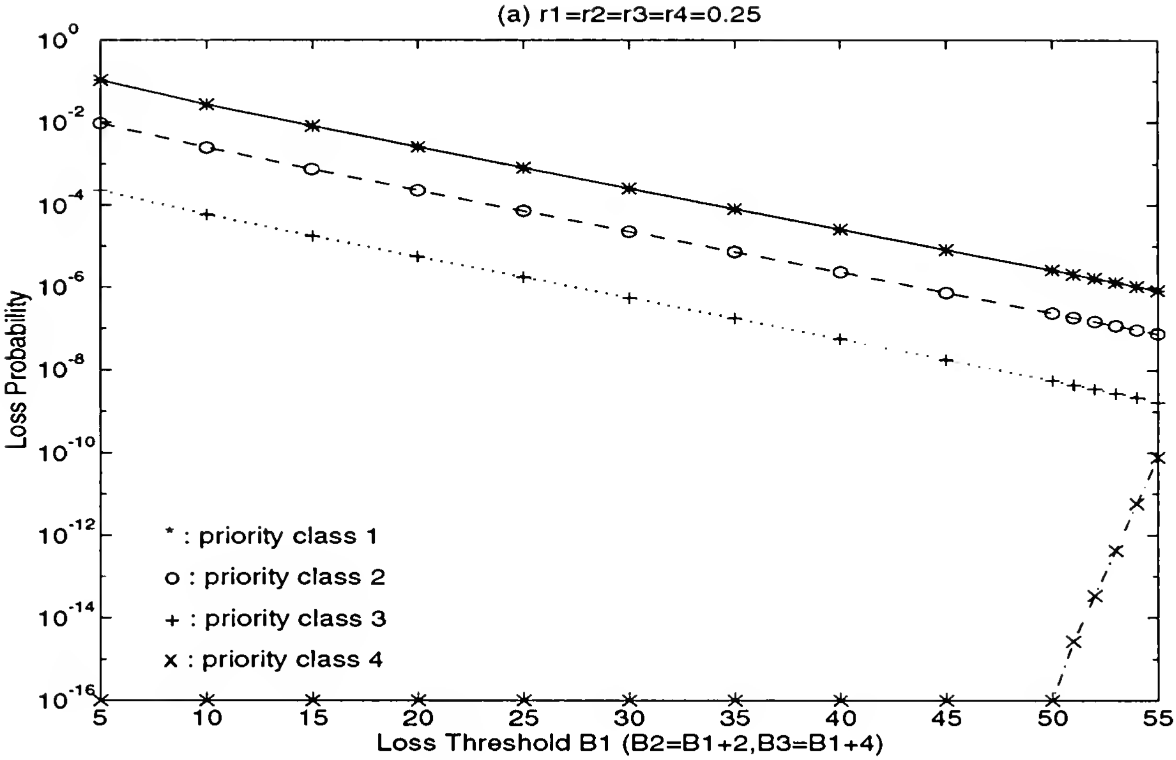
traffic mix ($r_1 = r_2 = r_3 = r_4 = 0.25$) and Figure 5.2(b) and Figure 5.2(c) show the results of two extreme situations, class 1 dominated ($r_1 = 0.7, r_2 = r_3 = r_4 = 0.1$) and class 4 dominated ($r_1 = r_2 = r_3 = 0.1, r_4 = 0.7$), respectively. Due to the limitation of the computing precision in our machine, the loss probabilities below 10^{-15} are offset by round-off errors and cannot be computed correctly. For those cases, a small number 10^{-16} will be substituted for them. It is noted that using a logarithmic scale, the loss probabilities display a linear-like variation as the threshold levels are increased, which is similar to the results of the binary priority system [70, 72, 73]. As the vthreshold levels increased linearly, the accessible buffer space for the associated priority class is expanded, thus reducing the loss probabilities. On the other hand, the increment of the threshold levels changes the state equilibrium probabilities. The buffer has a higher occupancy due to the increment, thus increasing the loss probability of the highest class. The curves giving the loss probabilities of the lower three priority classes remain parallel as the thresholds vary. Moreover, it is observed that the differences between the loss probabilities of the priority classes depend strongly on the traffic mixture.

Impact of traffic load. Next, the influence of different traffic loads under a balanced traffic mix is evaluated. The resulting loss probabilities are presented in Figure 5.3 with traffic loads of $\rho = 0.8$ and $\rho = 0.99$. It is seen that the variation of traffic load will effect the slopes of the curves giving the loss probabilities. With a set of fixed threshold levels, the increase of system load will rise the loss probabilities of all priority classes. In addition, it is observed that the loss probability of the most significant class can be improved substantially by sacrificing a moderate degree of the loss probabilities of lower priority classes.

Impact of buffer capacity. The variation of the loss probabilities as a function of buffer capacity is evaluated. Figure 5.4 shows the probabilities against various buffer capacities with a fixed traffic load $\rho = 0.9$ and different traffic mixes. The threshold levels are stabilized at a determined ratio of total buffer capacity ($B_1 = 0.7B, B_2 = 0.8B, B_3 = 0.9B$). All the loss probabilities decrease as the buffer capacity is expanded. Each curve of the loss priority classes varies with a different slope since the increments of the threshold levels are now proportional to the associated ratio, instead of a fixed distance. It is noticed that the variation of traffic mix has a substantial impact on the loss probabilities of the higher priority classes.

5.4 Optimization of Loss Thresholds

We now proceed to the problem of searching for a set of optimal loss thresholds within a finite buffer. The objective is to maximize the system admissible load without violating the given constraints of loss probabilities and traffic conditions. As mentioned earlier, under a fixed load condition, an increase in the loss threshold of a particular class can decrease its corresponding loss probability because the accessible buffer space for this class is increased. Simultaneously, the increase of the loss threshold will increase the loss probabilities of higher priority classes because the state equilibrium probabilities are changed. In other words, given a set of loss thresholds that satisfies the loss constraints, an increase in the loss threshold of a particular class will drive the loss probability away from the loss constraint of this class, while making the loss probabilities of higher priority classes move closer (or even exceed) to their loss constraints. Thus if there is any priority class for which loss probability violates its loss constraint, we can try to increase the corresponding loss threshold to reduce the loss probability (at the expense of increasing the loss probabilities of other classes). On the other hand, considering the effect of increasing system load under a set of fixed loss thresholds, the increase will raise the loss probabilities of all classes



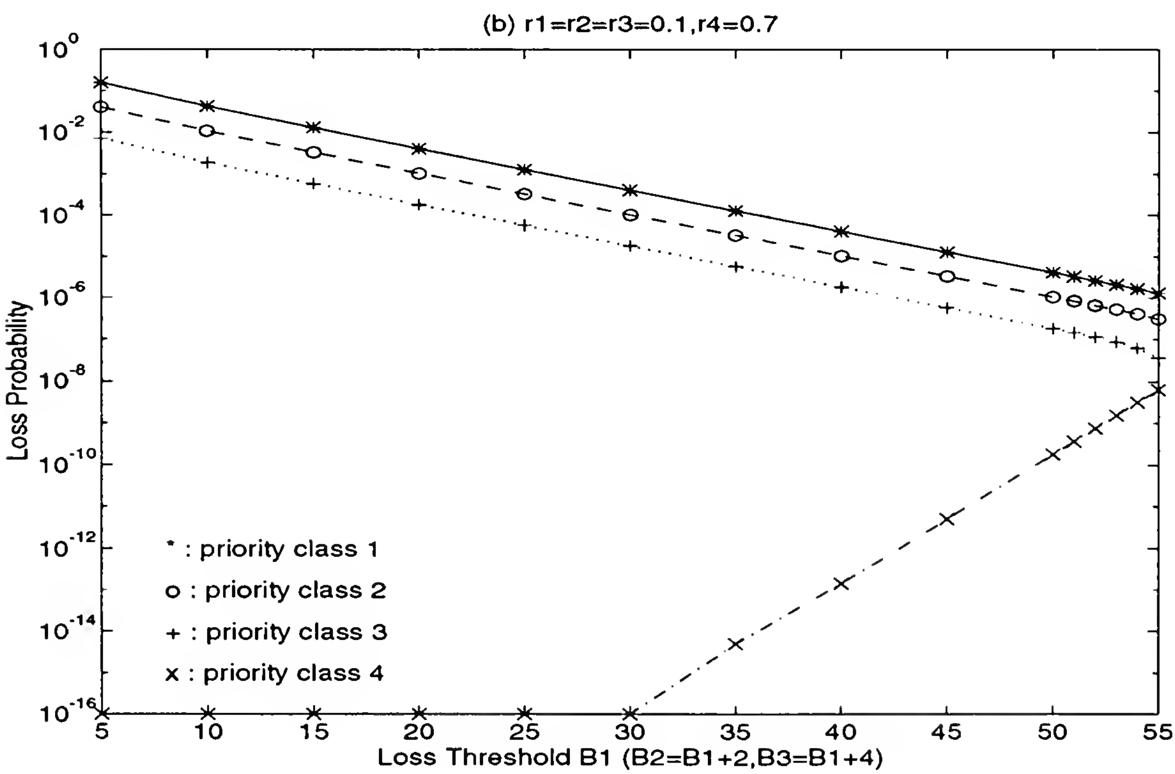


Figure 5.2: The Loss Probabilities as a Function of Threshold Levels with B=60 and $\rho = 0.9$ (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated

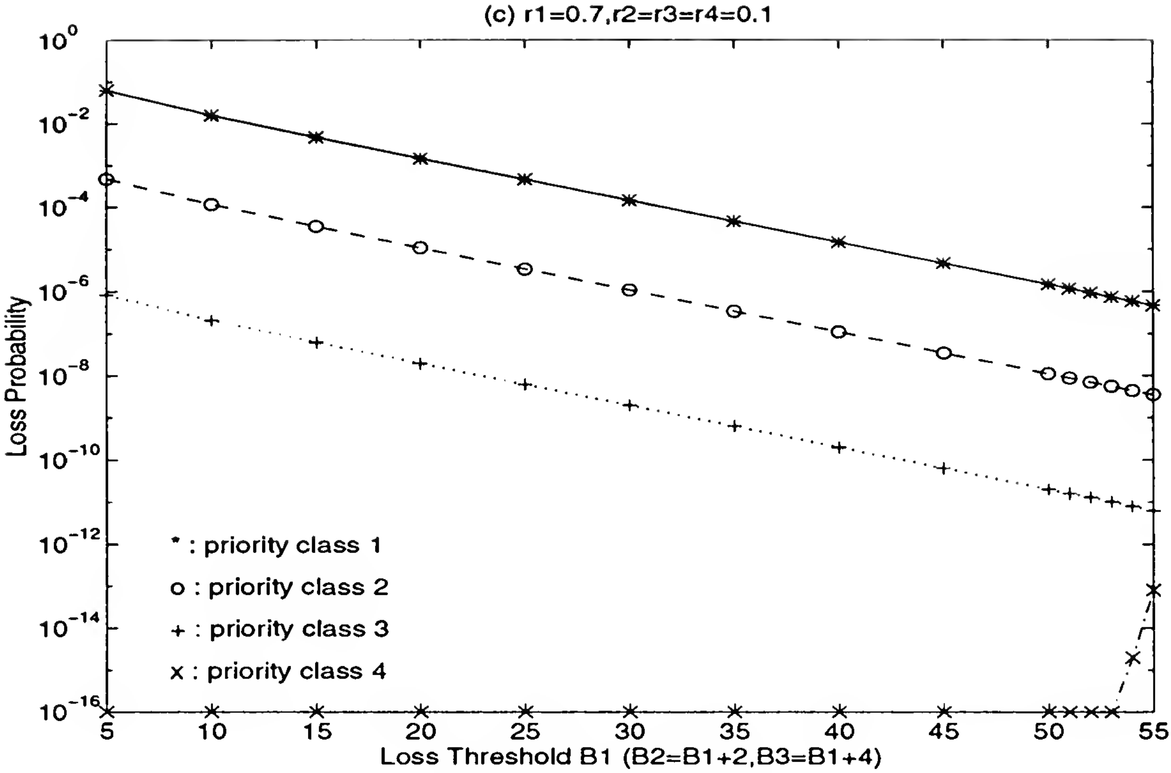


Figure 5.2--Continued

because the buffer state has higher occupancy. If we keep increasing the system load, eventually some of the loss constraints will be violated. Therefore, we can expect that the more rigorous the loss constraints, the lower the system admissible load.

Since we cannot change the given loss constraints and traffic conditions, we can only vary the loss thresholds for searching the optimal loss thresholds that maximize system admissible load. To make the optimization process more efficient, a procedure whose fundamental concept was introduced by Petr and Frost [69] is adapted. In their study, it has been verified, by exhaustive search, that the procedure always generates the optimal results for a buffer capacity up to 30. A similar algorithm is used to search for a set of thresholds which produces the best system admissible load under the given conditions. In addition, a well-established optimization technique, the Hooke and Jeeves optimization algorithm [71], is utilized for verifying the results and making fine adjustments.

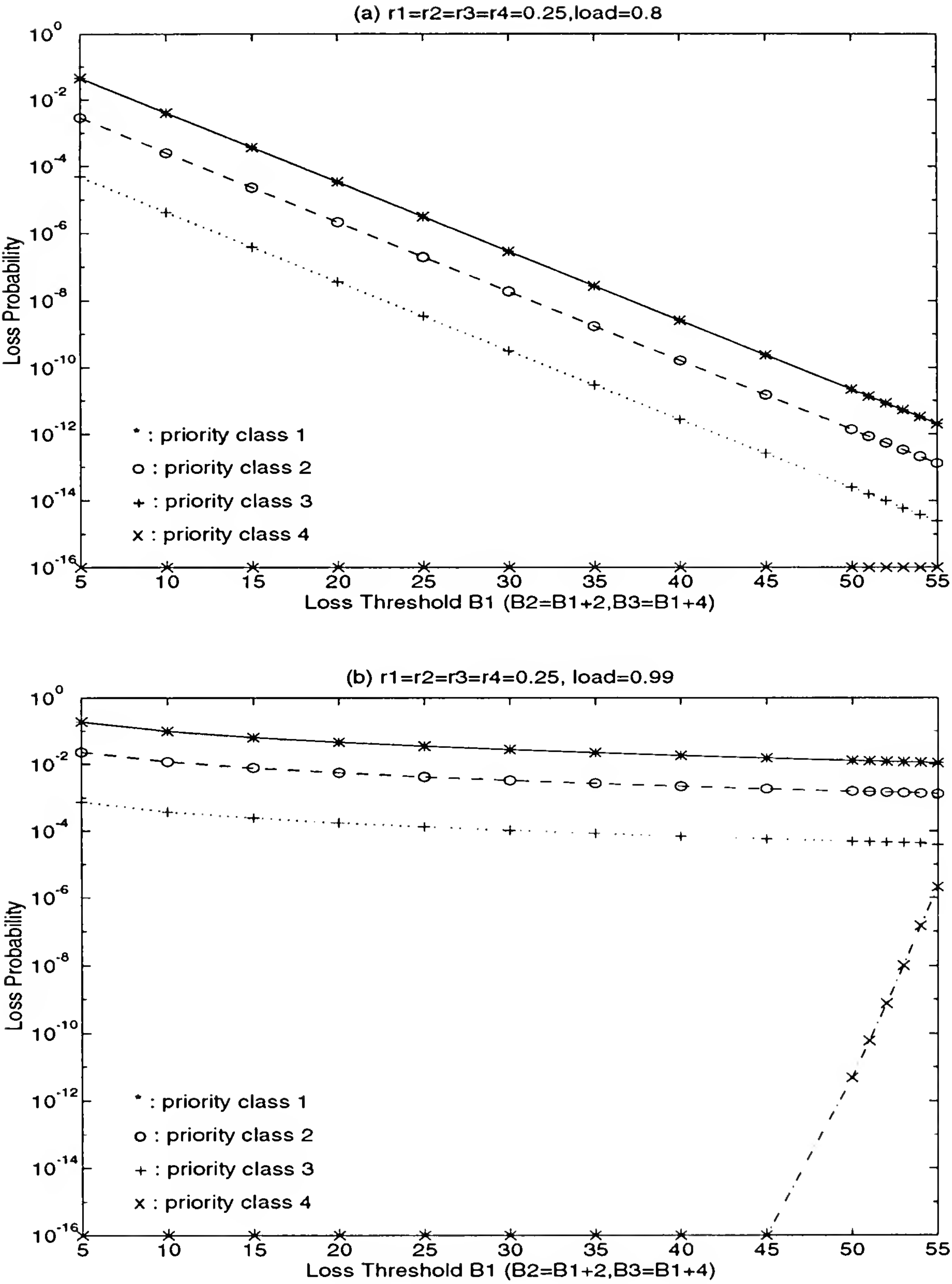


Figure 5.3: The Loss Probabilities as a Function of Threshold Levels with B=60 and Varied Traffic Load (a) $\rho = 0.8$ (b) $\rho = 0.99$

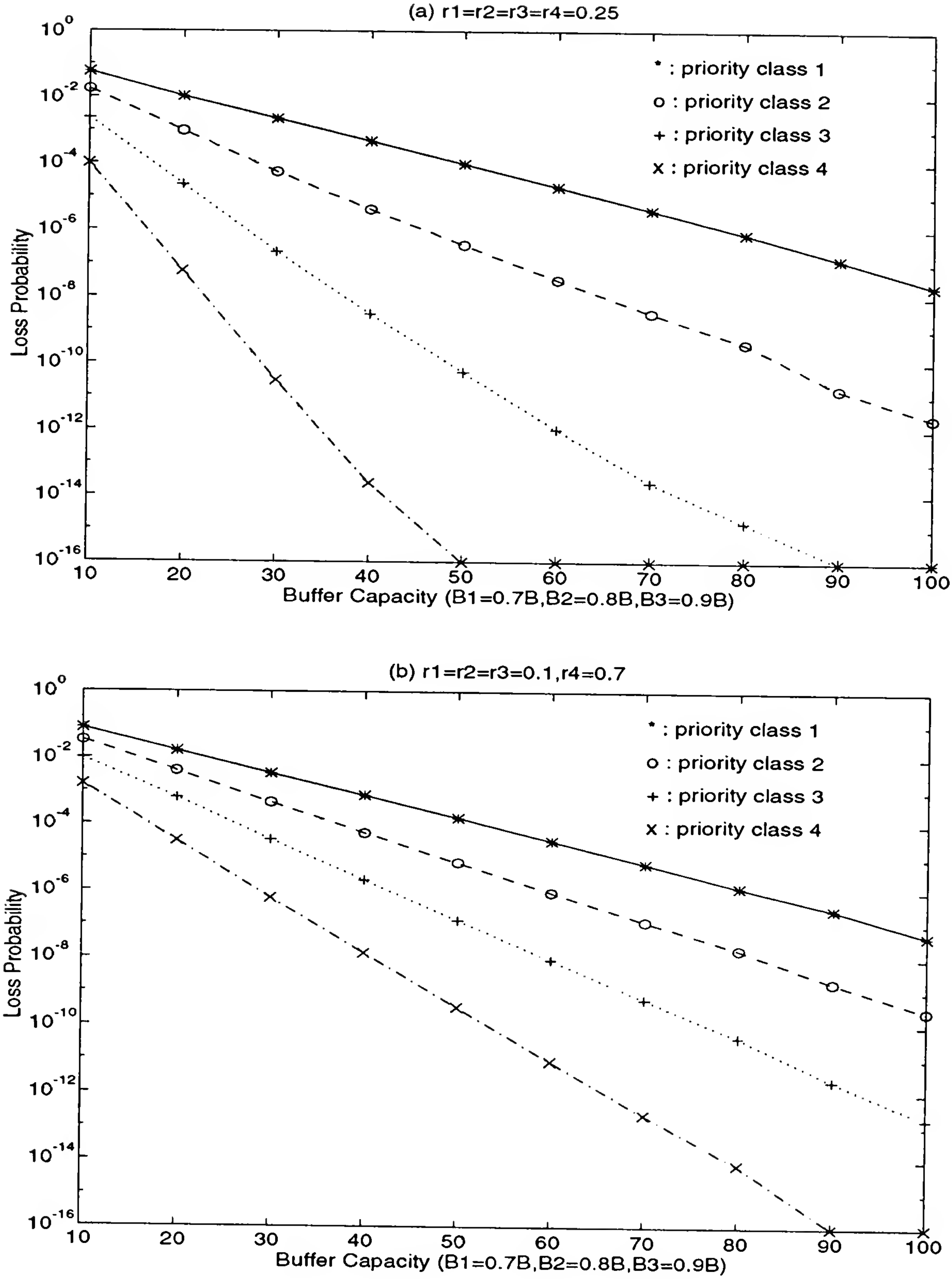


Figure 5.4: The Loss Probabilities as a Function of Buffer Capacity with $B_1 = 0.7B, B_2 = 0.8B, B_3 = 0.9B$ and $\rho = 0.9$ (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated

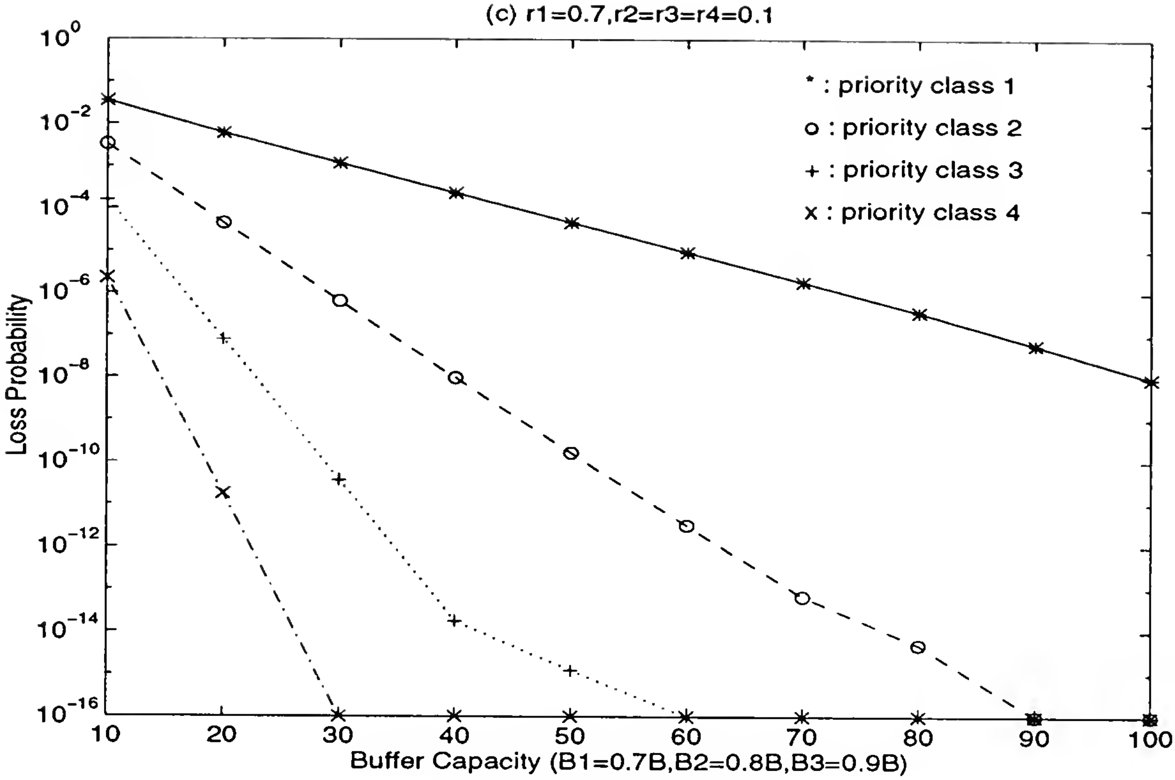


Figure 5.4—Continued

Because of the discrete nature in this problem, it is very difficult to prove by a rigorous mathematic means that the optimization procedures will always find a global optimum. In numerical optimization techniques, two standard heuristics that are frequently utilized to enhance the reliability of results are: (1) perturb the extremum by taking a finite amplitude step away from it and then see if the procedure leads to other better points or the same one, and (2) find local extrema starting from widely ranged initial base points and then pick the most extreme of these [71]. By making use of the heuristics as well as the exhaustive search method to run many tests on the results, we believe that employing the searching procedure and the Hooke and Jeeves method can always converge the system admissible load to an optimal point.

First the searching procedure is explained briefly. The searching procedure is initialized at a conservative base point with all the loss thresholds set to their

minimum values and ρ beginning at a small number. The loss probabilities at the base point are then evaluated to see if there is any priority class violating its loss constraint. If not, the base point is saved as a qualified base point and the ρ is increased by a prescribed step size to explore if higher ρ is possible. This procedure is repeated until some loss constraint is violated. At this point, we begin to search for a set of loss thresholds that satisfies the loss constraints under the current ρ. The corresponding loss threshold is increased by 1 each time to see if the reduced loss probability, as well as the loss probabilities of other classes, fulfill the requirements. If the exploration succeeds, the current ρ and loss thresholds become new base point and we try to increase ρ again. If it fails, then the previous base point is the highest ρ that can be achieved. Specifically, the searching procedure is terminated whenever the loss thresholds have reached their maximum values or the loss constraint of the highest class has been violated. The searching procedure, directing its moves based on the knowledge of the effects of changing system variables, has climbing property and can explore the searching space in a very efficient way to find the maximal system admissible load. The logic diagram for this searching procedure is presented in Figure 5.5.

The outcome of the above procedure is substituted into an optimization algorithm, the Hooke and Jeeves pattern search method, for verification and fine adjustment. The Hooke and Jeeves optimization algorithm is a direct search method that does not require the use of derivatives. The algorithm has ridge following properties and is based on the premise that any set of design moves that have been successful during early experiments is likely again to prove successful. The algorithm begins with an initial base point in the feasible design space and prescribed exploration step sizes. An exploration is then performed at a given increment along each of the independent-variable directions following the logic shown in Figure 5.6. In this case, the initial base point of the Hooke and Jeeves method is set to the final base point of

the searching procedure shown in Figure 5.5. For the sake of saving computing time, the step size in the searching procedure is chosen to be larger than the exploration step size in the Hooke and Jeeves algorithm (e.g., step size = 0.01 and exploration step size = 0.001) so that the outcomes can also be fine adjusted by the Hooke and Jeeves method. In addition, the Hooke and Jeeves algorithm can be evoked to adjust the system variables for an optimal system performance whenever the traffic conditions are changed. The previous base point can be used as an initial point with the selection of an appropriate exploration step size to search for the new optimal threshold arrangement in the dynamic environment.

The computational demand of the optimization process fluctuates from case to case. It is mainly dominated by the cost of evaluating loss probabilities under given loss thresholds and the size of the searching space. Since part of the state transition probabilities of the queueing model might need to be solved in a numerical way, the computation of loss probabilities can be demanding if the distance between loss thresholds is small. Fortunately, by employing the searching procedure, we are able to explore the searching space efficiently without making an exhaustive search over all of the possibilities. It has been shown that the search efficiency can be improved considerably for large N and buffer capacity, as compared to the exhaustive search method [69].

5.4.1 Numerical Examples: A Three-Class System

To demonstrate the effect of varying threshold levels, the system admissible load of a three-class system for all possible combinations of the threshold levels is computed. By simulating the three-class (two-threshold) system, we are able to display the results in a 3-D visualization. The buffer capacity is assumed to be 60 and $0 < B_1 < B_2 < B$. It is assumed that the loss constraints of the three classes are $PL_1 = 10^{-2}, PL_2 = 10^{-8}$ and $PL_3 = 10^{-14}$, respectively, and a balanced traffic mix is simulated ($r_1 = r_2 = r_3 = 1/3$). The maximum system admissible load for fixed

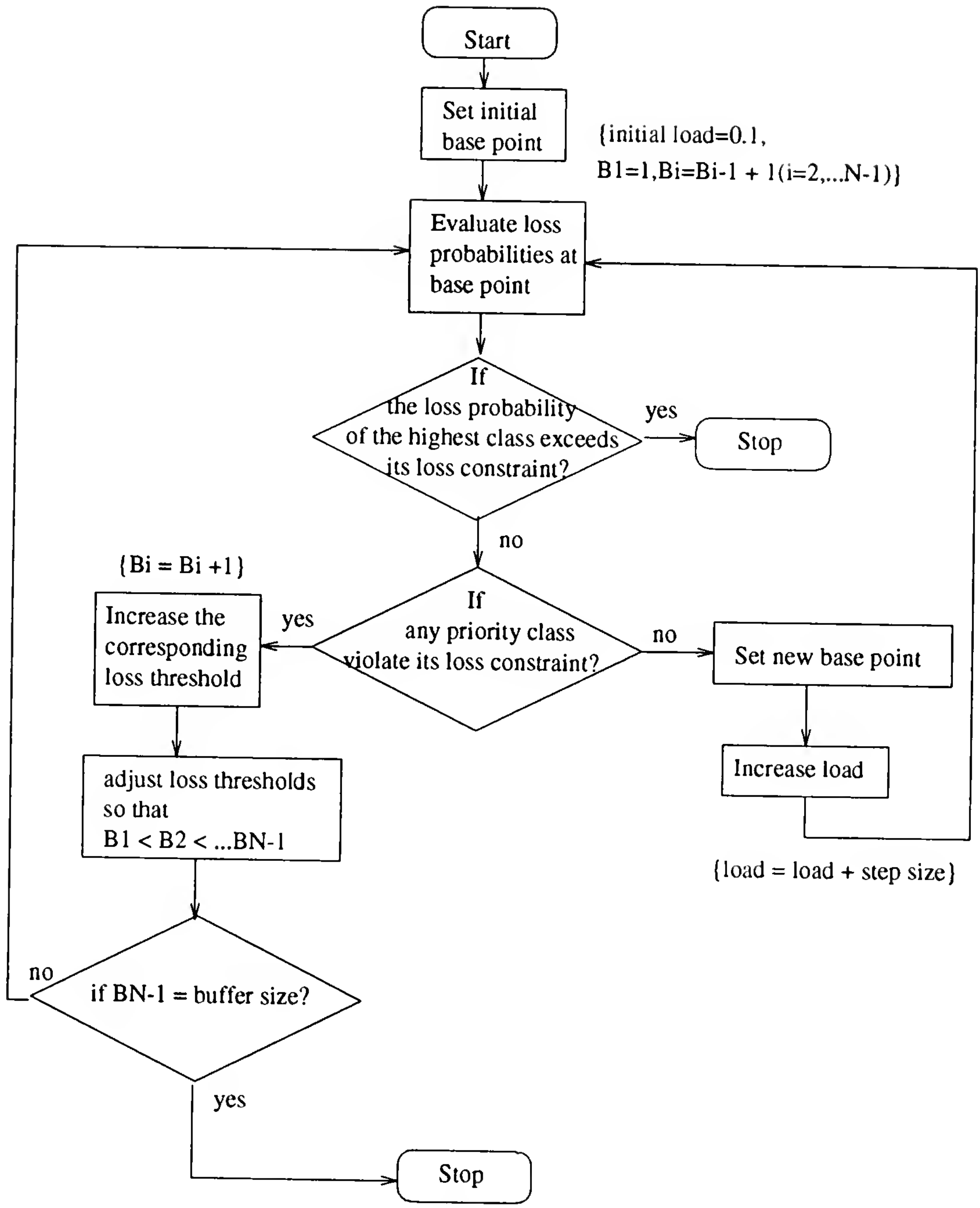


Figure 5.5: The Logic Diagram of the Searching Procedure

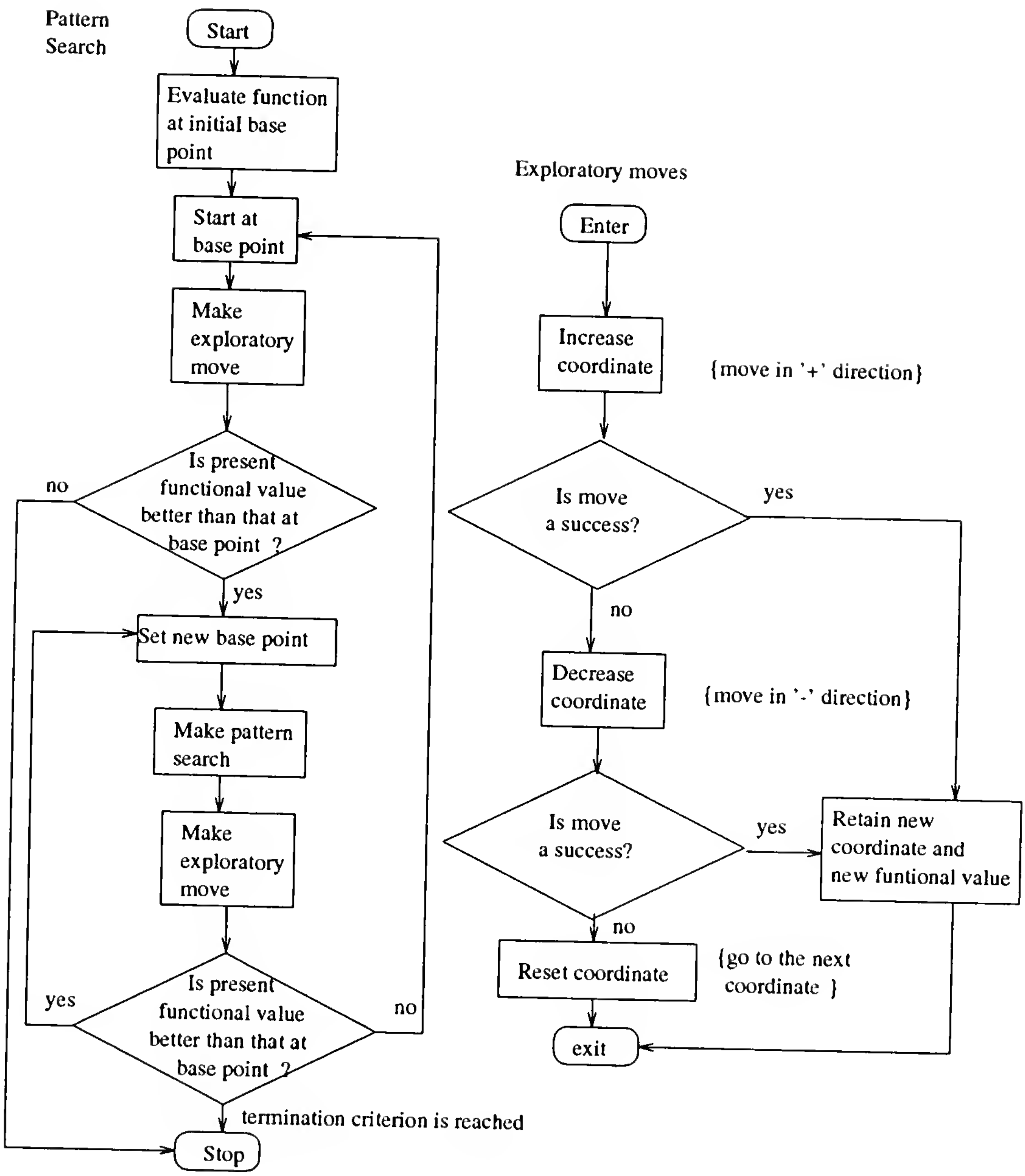


Figure 5.6: The Hooke and Jeeves Pattern Search Process

Table 5.1: The Optimal Loss Thresholds of a Three-class System with Balanced Traffic Mix ($PL_1 = 10^{-2}, PL_2 = 10^{-8}, PL_3 = 10^{-14}$)

B	B_1	B_2	ρ_{max}
10	3	7	0.294
20	6	15	0.721
30	12	24	0.894
40	20	34	0.953
50	27	41	0.971
60	39	54	0.987
70	46	63	0.992
80	58	73	0.996
90	59	74	0.997
100	72	89	1.0

threshold levels was computed by increasing the system load gradually until some of the loss constraints were violated. Figure 5.7 shows the maximum system loads that can be achieved for different combinations of the threshold levels B_1 and B_2. The optimal point in this case is shown to be $\rho = 0.987$ with $B_1 = 39$ and $B_2 = 54$. The searching procedure was utilized for finding the best threshold levels under the given conditions and it returned the values of $\rho = 0.98$ with $B_1 = 38$ and $B_2 = 52$. The outcome was then substituted into the Hooke and Jeeves algorithm and it was adjusted to $\rho = 0.987$ with $B_1 = 39$ and $B_2 = 54$, which is exactly the optimal point shown in Figure 5.7. The optimal system load and loss thresholds with respect to different buffer capacities are summarized in Table 5.1.

Next we demonstrate the potential performance improvement with the optimization procedures. Consider a three-class system with a set of loss constraints $PL_1 = 10^{-6}, PL_2 = 10^{-10}$ and $PL_3 = 10^{-14}$. We compare the maximal admissible loads of the systems with a fixed threshold assignment and with a set of optimal threshold levels. The buffer capacity is assumed to be 60 and a balanced traffic mix is simulated. The assignment of fixed threshold levels is arbitrary and an obvious choice is to select the levels which equally divide the available buffer space. Figure 5.8 and Table 5.2 present a comparison between the achievable loads with and without

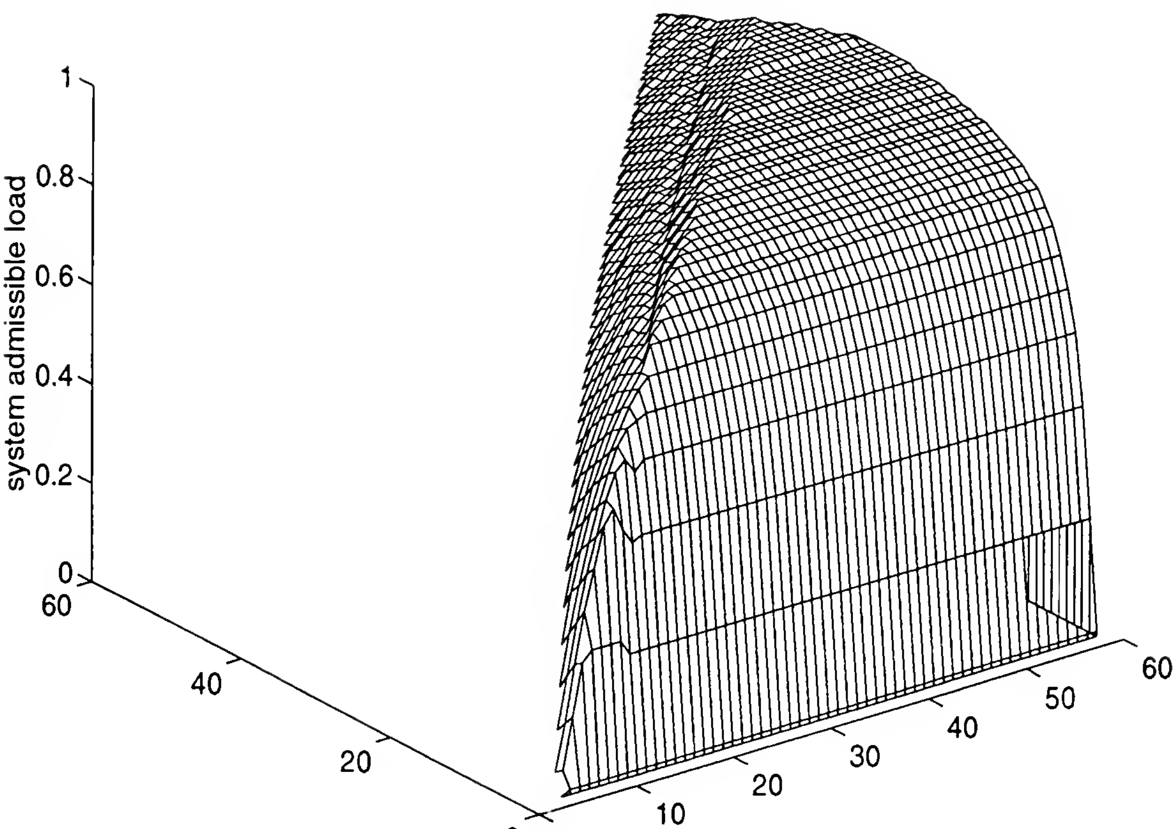


Figure 5.7: The System Admissible Loads as a function of Loss Thresholds ($PL_1 = 10^{-2}, PL_2 = 10^{-8}, PL_3 = 10^{-14}$)

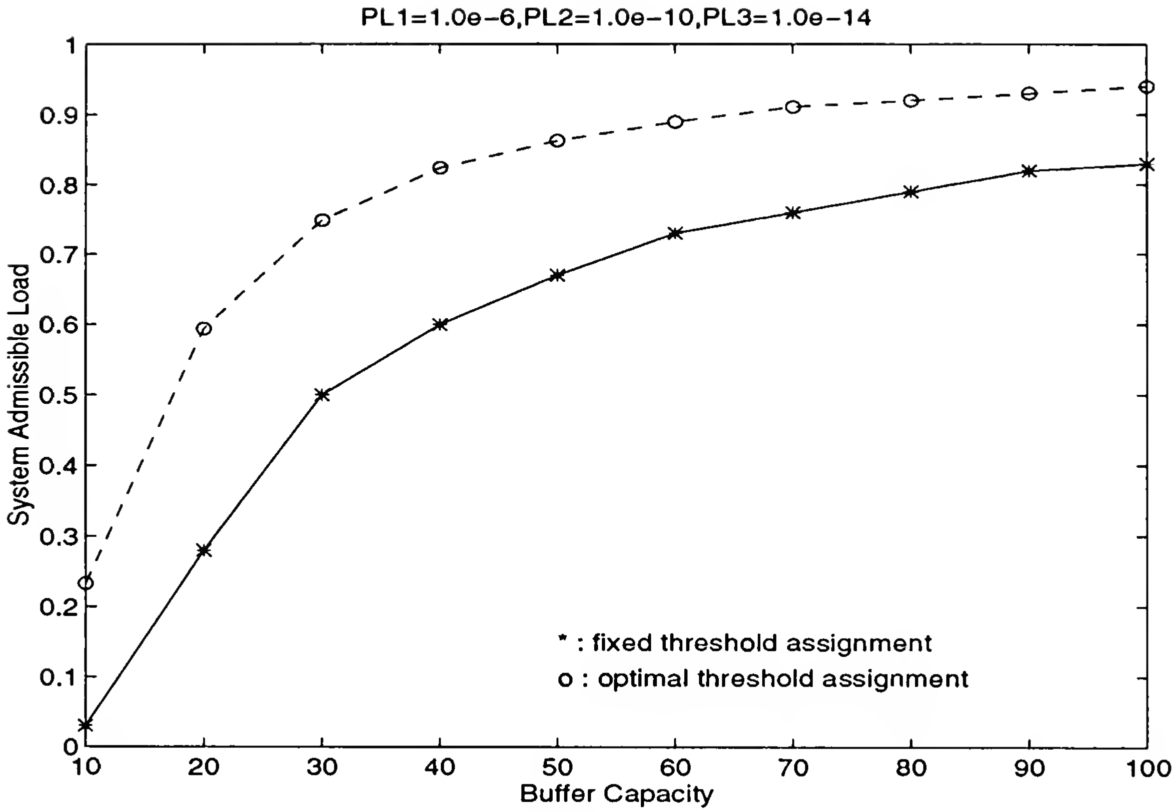


Figure 5.8: The System Performance Improvement with optimization ($PL_1 = 10^{-6}, PL_2 = 10^{-10}, PL_3 = 10^{-14}$)

optimization. It is seen that in this case the system can be improved up to 30% of the admissible load.

5.4.2 Numerical Examples: A Four-Class System

In this section the numerical results of a four-class system are presented. The numerical results could be used as a guideline to dimension optimally the finite buffer system at an ATM switching node. The optimal system admissible loads are computed using the procedures described above. Four sets of loss constraints were simulated: (A) $PL_1 = 10^{-2}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$, (B) $PL_1 = 10^{-4}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$, (C) $PL_1 = 10^{-6}, PL_2 = 10^{-8}, PL_3 = 10^{-10}, PL_4 = 10^{-12}$ and (D) $PL_1 = 10^{-8}, PL_2 = 10^{-10}, PL_3 = 10^{-12}, PL_4 = 10^{-14}$. In each case, three varied traffic mixes are simulated as in the previous numerical

Table 5.2: The Threshold levels and System Admissible loads with and without optimization ($PL_1 = 10^{-6}, PL_2 = 10^{-10}, PL_3 = 10^{-14}$)

	Fixed	T.		Optimal	T.	
B	B_1	B_2	ρ_{max}	B_1	B_2	ρ_{max}
10	3	6	0.03	6	8	0.233
20	6	12	0.28	12	17	0.594
30	10	20	0.50	20	26	0.749
40	10	26	0.60	29	36	0.824
50	16	32	0.67	38	45	0.863
60	20	40	0.73	47	55	0.89
70	23	46	0.76	58	66	0.911
80	26	52	0.79	65	73	0.92
90	30	60	0.82	75	83	0.93
100	33	66	0.83	87	95	0.94

examples. The maximum system admissible loads for each case are shown in Figure 5.9 and 5.10 as a function of buffer capacity and the optimal loss thresholds are presented in Table 5.3, 5.4, 5.5 and 5.6 respectively. It is seen that the system with the most rigorous loss constraints (i.e. case D) has the lowest optimal system admissible load as expected. Furthermore, we see that in all cases, the system with class 1 (the lowest priority class) traffic dominating, displays the best admissible load, while the system with class 4 (the highest priority class) traffic dominating, has the poorest system performance.

In case A a system with widely spread loss constraints is considered, each priority class having a loss constraint with 4 orders of magnitude difference between adjacent classes. It is noticed that the system admissible load is sensitive to the variation of traffic mixture. The diversity of the admissible load can be more than 30% when the buffer capacity is small. Following the case a system with similar loss constraints, but with decreased loss limits of class 1 by two orders of magnitude is simulated. The results indicate a minor degrade of the admissible load when class 4 traffic dominated while a reduction of about 10% in the load (with a small buffer capacity) when class 1 traffic dominated. In case C, a system is simulated with

Table 5.3: The Optimal Loss Thresholds with Different Traffic Conditions (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated ($PL_1 = 10^{-2}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$)

	(a)				(b)				(c)			
B	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}
10	3	6	8	0.269	2	5	9	0.181	3	6	8	0.463
20	7	13	17	0.711	4	10	17	0.531	11	15	19	0.881
30	12	21	26	0.875	7	16	24	0.711	19	23	26	0.952
40	18	29	34	0.933	10	23	33	0.812	25	29	32	0.97
50	30	41	47	0.966	14	31	42	0.87	40	44	47	0.99
60	32	43	49	0.971	17	36	49	0.9	40	44	47	0.99
70	47	58	64	0.982	24	46	60	0.931	40	44	47	0.99
80	47	58	64	0.982	30	55	70	0.95	64	68	71	1.0
90	68	81	87	0.991	36	64	79	0.96	64	68	71	1.0
100	69	81	87	0.994	45	74	90	0.97	64	68	71	1.0

a smaller difference between the loss constraints, each priority class having a loss constraint with two orders of magnitude difference between adjacent classes. It is seen that the optimal system admissible load displays only a moderate change as the traffic mix varies. Finally, the results of case D were obtained by decreasing each loss constraint that is assumed in case C by a two orders of magnitude. The results show a drop of about 10% in the load when the buffer size is small, while a minor degrade occurs when the buffer size is large. Based on these results, we can conclude that no matter what loss constraints and traffic conditions are, the system performance can always be improved to a reasonable level (e.g., 0.9) by dimensioning the buffer capacity to a moderate size and imposing proper threshold levels.

5.5 Conclusion

In this chapter, a queueing model to analytically characterize a multiple-priority shared buffer system controlled by the PBS scheme is presented. The queueing system is modeled as a discrete time Markov chain with finite buffer capacity B and N classes of arrivals. With the assumption of a multinomial traffic distribution, an queueing model which predicts accurately the steady-state loss probabilities of a

Table 5.4: The Optimal Loss Thresholds with Different Traffic Conditions (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated ($PL_1 = 10^{-4}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$)

	(a)				(b)				(c)			
B	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}
10	4	6	8	0.25	4	5	9	0.171	5	6	9	0.331
20	10	13	18	0.641	8	11	16	0.521	13	15	19	0.731
30	19	22	27	0.8	12	17	24	0.693	22	24	27	0.85
40	28	32	37	0.873	19	24	33	0.781	32	34	37	0.901
50	35	39	45	0.9	24	31	42	0.842	39	41	44	0.921
60	43	47	53	0.92	32	40	52	0.88	50	52	55	0.94
70	57	61	67	0.94	38	47	60	0.902	59	61	64	0.95
80	66	71	77	0.95	47	57	70	0.920	71	73	76	0.96
90	66	72	77	0.95	53	63	73	0.931	71	73	76	0.96
100	80	85	91	0.96	61	72	86	0.94	90	92	96	0.97

Table 5.5: The Optimal Loss Thresholds with Different Traffic Conditions (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated ($PL_1 = 10^{-6}, PL_2 = 10^{-8}, PL_3 = 10^{-10}, PL_4 = 10^{-12}$)

	(a)				(b)				(c)			
B	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}
10	6	8	9	0.27	6	8	9	0.233	7	8	9	0.318
20	14	17	19	0.62	12	15	18	0.56	16	18	19	0.673
30	23	27	29	0.76	19	24	27	0.705	25	27	29	0.788
40	31	35	37	0.821	26	32	36	0.781	33	35	37	0.84
50	40	44	46	0.86	34	40	45	0.83	44	46	48	0.881
60	51	55	58	0.892	41	48	54	0.86	52	54	56	0.90
70	61	65	68	0.91	49	56	62	0.88	64	66	68	0.92
80	68	72	75	0.92	58	66	72	0.90	73	75	77	0.93
90	78	83	85	0.93	64	73	79	0.91	84	86	88	0.94
100	91	96	98	0.94	72	81	88	0.92	84	86	88	0.94

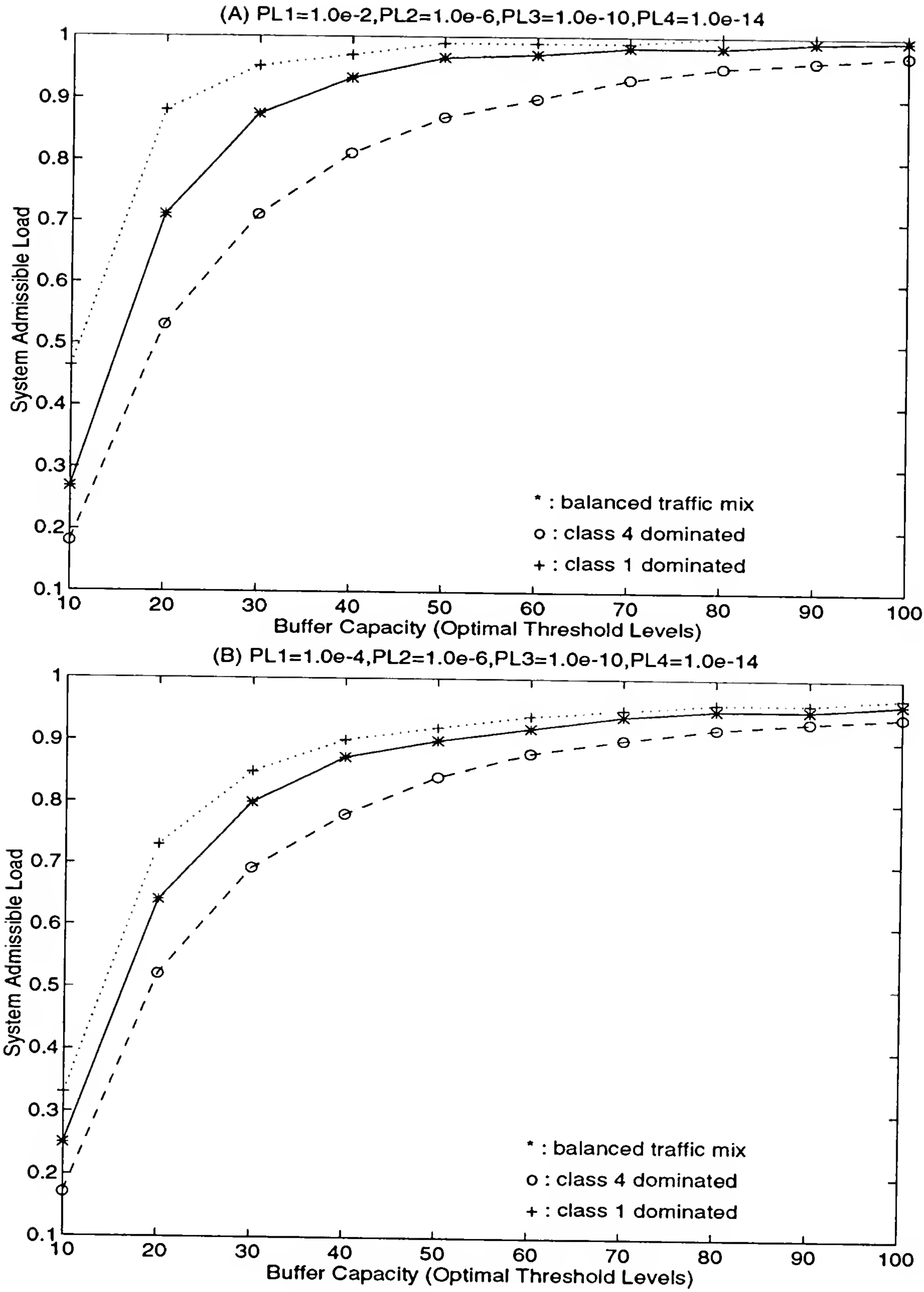


Figure 5.9: The System Admissible Loads as a function of Buffer Capacity (A) $PL_1 = 10^{-2}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$, (B) $PL_1 = 10^{-4}, PL_2 = 10^{-6}, PL_3 = 10^{-10}, PL_4 = 10^{-14}$

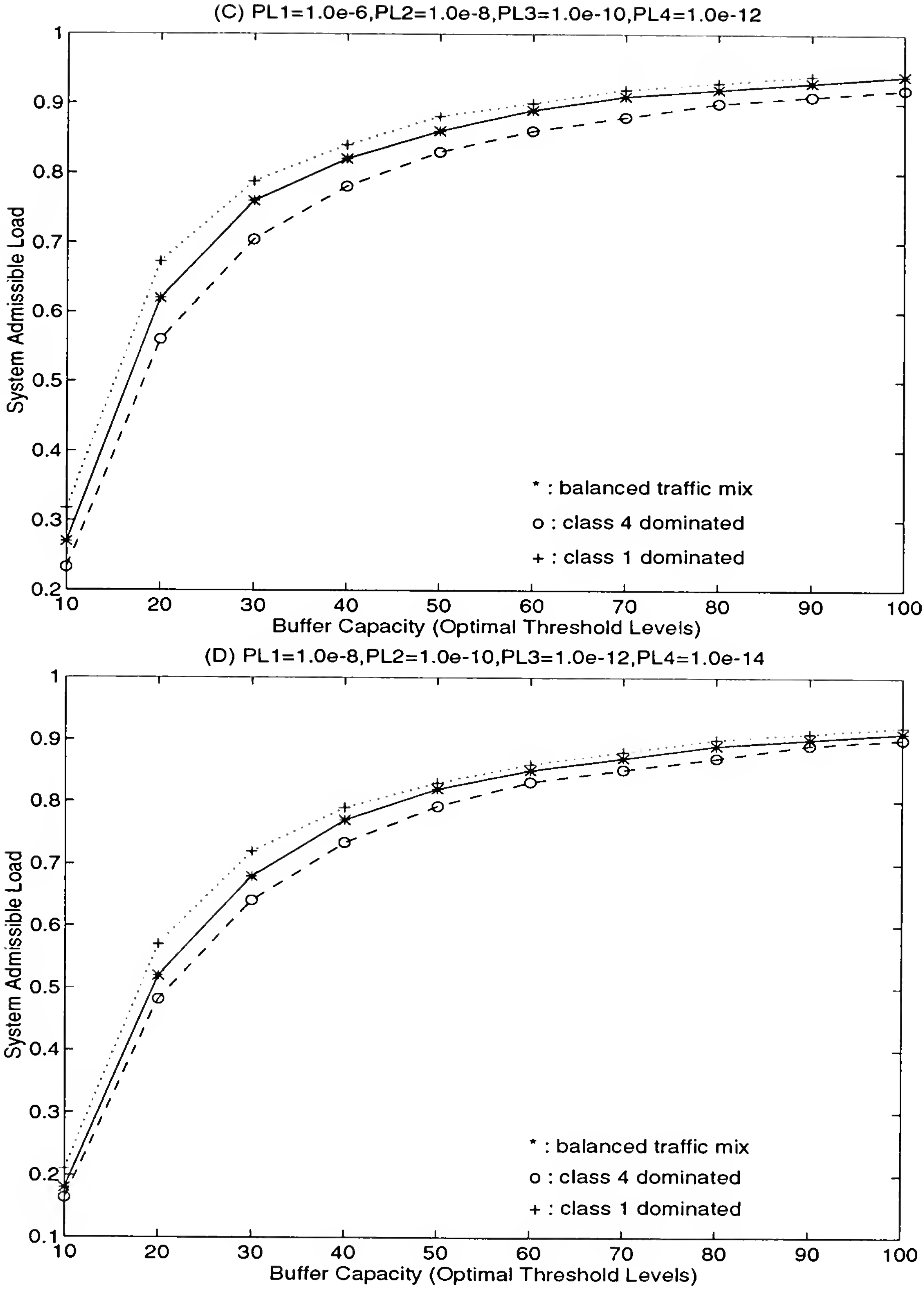


Figure 5.10: The System Admissible Loads as a function of Buffer Capacity (C) $PL_1 = 10^{-6}, PL_2 = 10^{-8}, PL_3 = 10^{-10}, PL_4 = 10^{-12}$, (D) $PL_1 = 10^{-8}, PL_2 = 10^{-10}, PL_3 = 10^{-12}, PL_4 = 10^{-14}$

Table 5.6: The Optimal Loss Thresholds with Different Traffic Conditions (a) Balanced Traffic Mix (b) Class 4 dominated (c) Class 1 dominated ($PL_1 = 10^{-8}, PL_2 = 10^{-10}, PL_3 = 10^{-12}, PL_4 = 10^{-14}$)

	(a)				(b)				(c)			
B	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}	B_1	B_2	B_3	ρ_{max}
10	7	8	9	0.188	7	8	9	0.164	7	8	9	0.21
20	15	17	19	0.529	13	16	18	0.482	16	18	19	0.57
30	23	26	28	0.684	21	24	29	0.641	26	28	29	0.72
40	33	36	39	0.772	29	33	37	0.734	35	37	39	0.79
50	42	46	48	0.821	37	43	47	0.792	43	45	47	0.83
60	51	55	57	0.852	46	52	57	0.831	52	54	56	0.86
70	58	62	65	0.87	52	59	64	0.851	61	63	65	0.88
80	69	73	76	0.891	60	67	73	0.87	72	74	76	0.90
90	75	79	82	0.90	71	79	85	0.891	80	82	84	0.91
100	84	88	91	0.911	78	86	92	0.90	90	92	94	0.92

multiple-class system under a set of given loss thresholds is developed. Making use of the priority control, the system is able to support different grades of loss quality with a minimum hardware cost and also provides a better control on the delay introduced during buffering.

In order to realize the full potential of the PBS control mechanism, optimize system performance based on the given system criteria is crucial. A numerical searching procedure is introduced to find the best thresholds which produce the best admissible load, under the given conditions. A well-established optimization technique, the Hooke and Jeeves optimization algorithm, is utilized to verify the results and make fine adjustments. By employing the searching procedure, we are able to explore the searching space efficiently without making an exhaustive search over all possibilities. It has been shown that the system performance can be improved significantly by the optimization procedures. From the results, we see that the system with more rigorous loss constraints produces a lower optimal system admissible load. Moreover, it is observed that the system with the lowest priority traffic dominated displays the best admissible load, while the system with the highest priority traffic

dominating has the poorest system performance. It is noticed that the system admissible load is sensitive to the variation of traffic mixture, when buffer capacity is small and loss constraints are widely spread. We show that, with dimensioning the buffer capacity to a moderate size and imposing proper threshold levels, the system performance can always be improved to a reasonable level. The results provide a close insight into the PBS scheme to manage effectively the finite buffer system at ATM switching nodes.

CHAPTER 6
COMPARATIVE STUDY OF ATM FLOW CONTROL MECHANISMS

6.1 Congestion Control Mechanisms for Best-Effort Service

The issue of traffic management for the best-effort (or Available Bit Rate) service in ATM networks has raised lots of attention recently [56, 57, 58, 68]. The service class is designed to support the existing data networking applications, which are likely to be dominant at the initial phase of ATM networks. As mentioned in Chapter 4, this class of applications typically has unpredictable characteristics. In general, as a data application requests a communication channel via the host's operating system, neither the user nor the host's operating system is capable of predicting how the application will be used during the connection's lifetime. It is thus difficult to require these applications to specify their bandwidth requirements in advance of transmission so that the data transfer can be managed by strict admission control. As a consequence, the network is not able to provide any guarantee on the allocated network resource, as well as QOS for the applications.

The provision of ABR service in ATM networks is attractive for the following reasons. With the conservative resource allocation policy of the connection admission control to provide QOS guarantees for critical traffic, the ABR applications can improve network resource utilization by allowing multiple ABR connections to share dynamically the unused network bandwidth left over by the guaranteed traffic. In addition, the users need not declare bandwidth parameters other than the peak cell rate at the connection setup. The ABR applications are allowed to access to the maximum network bandwidth as long as it is available.

It is well recognized that additional flow control mechanisms must be applied to the ABR applications since there is no bandwidth control at traffic entry. Congestion may occur at critical network resources and induce an unacceptable network performance, if no adequate resource control is provided. It is the responsibility of a switching node to ensure that the QOS of the guaranteed traffic will not be adversely affected by the ABR traffic under overload condition. The most direct approach to solve this problem is to allocate a minimum cell buffering space for the guaranteed traffic and make the link scheduling server always service the guaranteed traffic with a higher priority than the ABR traffic. Although this policy provides resource protection for the guaranteed traffic, it cannot prevent or reduce cell loss of the ABR traffic. Many data applications are relatively endurable for delay but sensitive to cell loss. If a single cell of a packet is lost, the traffic source will have to retransmit the entire packet to recover from the information loss. The unnecessary retransmission of other successfully transmitted cells of the packet may cause sustained congestion and resource inefficiency.

A major portion of the existing data networking traffic, including the exponentially growing Internet traffic, is carried by using the Transmission Control Protocol (TCP). TCP is a transport layer protocol which provides users with reliable connection-oriented services across an unreliable network. TCP is heavily used in combination with Internet Protocol (IP), which supports the interconnection of networks by using a datagram service. Since TCP connections are expected to be one of the major applications of ABR service, the performance of sending TCP traffic over ATM networks needs to be considered carefully.

To carry TCP/IP traffic across ATM networks, the IP datagrams have to be fragmented, or *cellified*, into small fixed-size cells. It has been shown that due to the fragmentation process, the effective throughput (or goodput) of a TCP connection can be significantly reduced as compared with the packet (*non-cellified*) TCP during

network congestion [77]. The main reason for the low goodput of the *cellified* TCP is that when traffic bursts arrive at the switch, cells belonging to different packets are interleaved. Thus when the ATM switch buffer is overflowed and begins to discard the arriving cells, each discarded cell is likely to belong to a different packet. Those packets corrupted by cell loss will be retransmitted by the traffic sources after a timeout. However, all the other cells from such packets, which will be eventually discarded at the destinations, could still be transmitted over the network. Therefore a considerable amount of link bandwidth can be wasted sending the useless cells. While in the packet TCP, the phenomenon doesn't occur since the entire packet will be dropped as the buffer is full.

Another reason that contributes to the poor performance for sending TCP over ATM is the synchronization behavior of the TCP congestion control algorithm between different connections, which has been studied for packet TCP in [78, 79, 80]. TCP uses an adaptive window mechanism to control the traffic amount sent by the traffic source in response to network conditions. When a packet loss is detected by an expiration of a retransmission timer, the adaptive window size is reduced to one maximum size packet (i.e., maximum transfer unit or MTU) and will be incremented upon receiving positive acknowledgments of packets transmitted subsequently. The fragmentation process in ATM aggravates the possibility of cell dropping from many connections during congestion. As a consequence, those connections are forced to reduce their window to one in the same period and begin to transmit their packets in a synchronized pattern. The traffic synchronization decreases the link utilization since the small window size, which controls the traffic amount that can be transmitted during one round trip time, makes the traffic sources idle simultaneously for a significant time. Moreover, in the worst case, the large burstiness of the synchronized traffic induces another major cell loss, when a buffer doesn't have enough space to

accommodate the traffic and then the same packets are retransmitted repeatedly, driving the connections to a shutoff status.

One effective way to solve the high retransmission rate of TCP traffic over ATM is to use a feedback flow control mechanism that can force traffic sources to slow down when the network is congested. Under optimal circumstance, each virtual connection transmits with its fair share of the available bandwidth or with the maximum bandwidth if there is no other connection transmitting. Two classes of feedback control mechanisms have been proposed for this purpose: credit-based and rate-based. The credit-based mechanism (see Figure 6.1), which is performed on a link-by-link per-virtual-connection basis, is similar to the sliding window flow control scheme in existing data networks. The traffic amount that can be transmitted over a link is controlled by a credit count at the sending end of the link, where each unit of credit represents one empty cell buffer at the receiving end that is allocated to the connection. While with the rate-based control mechanism (see Figure 6.2), the traffic control is performed on a end-to-end per-virtual-connection basis. The congested node generates feedback signals (e.g., BECN cells) to regulate the rate at which each source emits cells into the network on every connection.

It has been shown that the credit-based approach offers excellent bandwidth utilization with zero cell loss in a high-loaded network [68, 81]. As congestion is encountered on a path, the buffer of the virtual connection at the congested node is filled up gradually due to the imbalance between input and output. This will force the transmission process of its upstream node to slow down and finally create an effect of link-by-link backpressure propagating back to the traffic source. Thus the traffic emission at the source can be throttled and the excessive traffic can be blocked at the edge of the network. With the control schemes, the network bandwidth can be fully utilized by sharing equally between competing users and packet retransmission can be prevented.

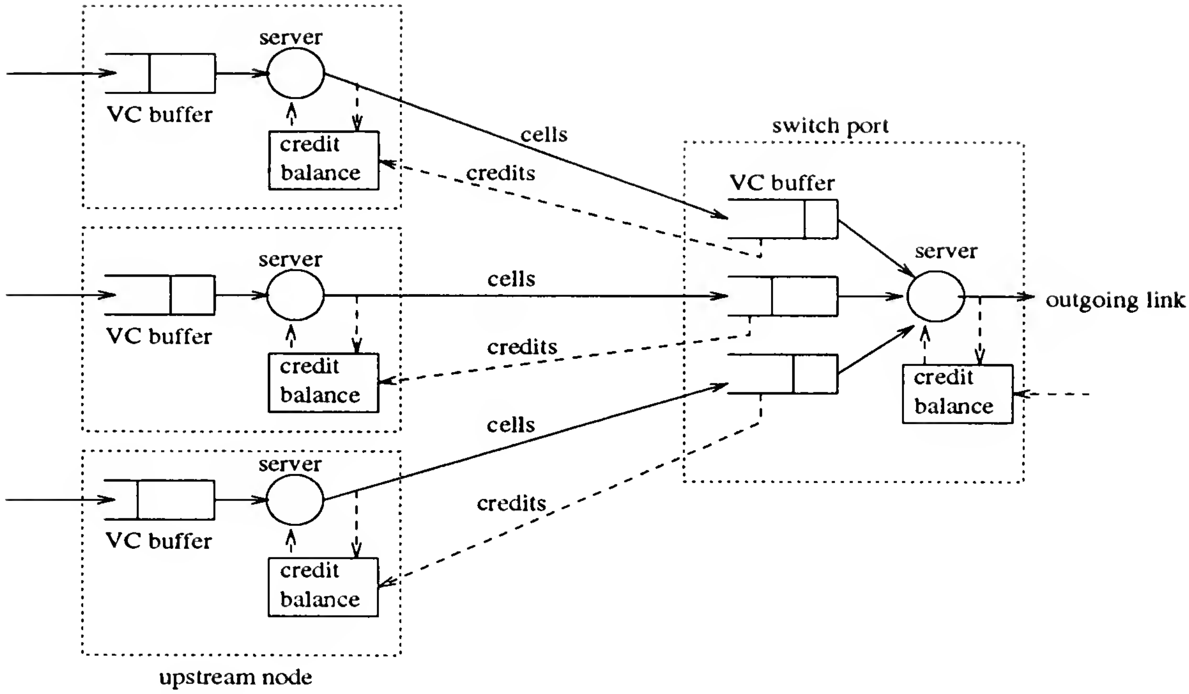


Figure 6.1: The link-by-link credit-based flow control mechanism

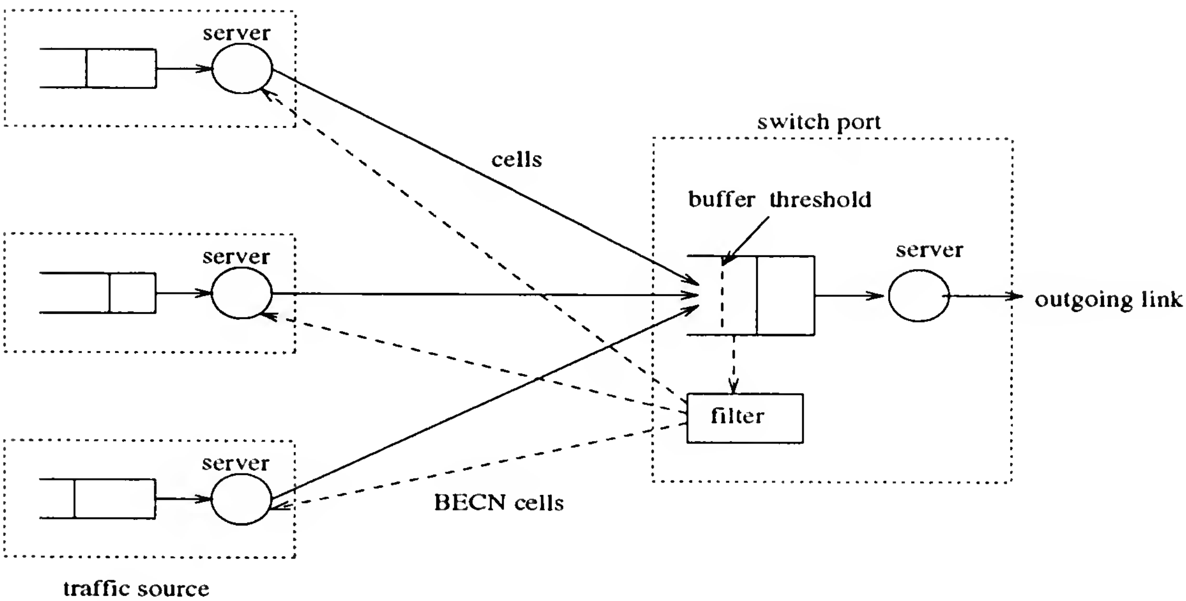


Figure 6.2: The rate-based flow control mechanism

However, the implementation of the credit-based control scheme is costly. It requires per-virtual-connection buffering and a counting process on each switch port. In addition, a scheduling server that can schedule cells for transmission on a per VC basis is needed. For a large ATM switch that supports thousands of connections, the complexity of implementation and computational burden are substantial. The credit-based approach also requires some fraction of the link bandwidth on the reverse direction of the connection to be reserved for the transport of credit cells. The buffer requirement of each connection is proportional to the targeted average bandwidth and round-trip link delay time.

Because of the significant complexity on switch hardware required by the credit-based approach, the network providers of public networks are likely to support a rate-based approach [57]. Unlike credit-based schemes, rate-based flow control doesn't require per-virtual-connection buffering and counting and it only reacts when the buffer occupancy exceeds a threshold. The rate-based approach thus imposes very moderate requirements on the switch hardware. On the other hand, rate-based schemes may not attain optimal link utilization and, as traffic is extremely bursty, it is possible for cell loss to occur. In addition, when the round-trip delay between the traffic source and the congested node is larger, the rate-based approach suffers a slower adjustment speed on the bandwidth of each connection since it takes a longer time for an explicit congestion indicator to travel back to the source.

Although rate-based control schemes have been described in some recent works [57, 82, 83], it is still not clear how the system parameters should be fine tuned to give better network performance. Moreover, the issues of connection transient behavior, fairness of resource sharing, and whether the scheme can adapt well to more complex network configurations haven't been investigated yet. This chapter gives a detailed analysis of a rate-based control scheme proposed in [83] and conducts a comparative study of the performance improvement for sending TCP over ATM

by applying the credit-based and rate-based flow control schemes. It has been shown that with well-tuned system parameters, the control schemes can prevent network throughput being severely affected by packet retransmissions and provide substantial performance improvement. In addition, a modified rate-based approach that can provide better network performance has been suggested and demonstrated.

6.2 The TCP Adaptive Window Algorithm

In this section a brief overview of the main features of the TCP adaptive window algorithm is presented to offer a better understanding of the behavior of TCP traffic. TCP provides users with reliable data transport by using positive acknowledgment in conjunction with retransmission. The TCP protocol defines a sliding window (denoted by *wnd* here) used by the senders to control traffic flows into network. The window size is adjusted dynamically in response to network condition and can never be increased to a larger size than the receiver advertised window, which reflects the buffering space at the receivers.

Whenever a packet loss is detected, a control threshold (denoted by *ssthresh*) is set to half of the current window and the window size is adjusted to one MTU. The window will be increased every time a positive acknowledgment of the packet transmitted subsequently is received. The window recovery has two phases: the slow start and congestion avoidance phases. As the window size is below the *ssthresh* (or in the slow start phase), the window increment is multiplicative, that is, the window size is doubled after an entire window's worth of packets have been acknowledged. Once the window size reaches the *ssthresh*, the window recovery process slows down and the congestion avoidance phase begins. The window is expanded by a small increment each time a packet is acknowledged, by which the window size will be increased by one MTU after a full window's worth of packets have been transmitted successfully.

The packet loss in TCP is detected by either the expiration of a sender's retransmission timer or the receipt of duplicate acknowledgments. The interval of the retransmission timeout (denoted by RTO) is chosen by taking into account an estimation of the average round trip time (RTT). TCP updates the estimated average RTT each time it obtains a new measurement to reflect the transient traffic condition in the network. The RTO is set equal to $\beta \times RTT$ (β equal to 2 in this study).

The following procedure [78] summarizes the TCP adaptive window algorithm described above.

- When a packet loss is detected, the sender does

  $ssthresh \quad = \mathrm{MAX}[wnd/2,\ 2];$
  $wnd \quad = 1;$

- When a new packet is acknowledged, the sender does

  if $\quad (wnd < ssthresh)$
  $\quad\quad wnd \mathrel{+}= 1;$
  else
  $\quad\quad wnd \mathrel{+}= 1 \,/\, \lfloor wnd \rfloor;$

A simulation model of traffic sources running with the TCP adaptive window algorithm has been developed. Without loss of generality, it is assumed that all window sizes are measured in units of MTU and a TCP session will always transmit the maximum size packets. Figure 6.3 shows the traces of the packet sequence number sent by a TCP connection and the corresponding window size. The maximum window size is assumed to be 8.

Consider a network configuration consisting of a single bottleneck link interconnecting two switches, as shown in Figure 6.4. Eight TCP connections are sending data traffic concurrently in the same direction through the ATM network. Assume that each connection uses a maximum window size of 8 (= 64 Kbytes) with a constant packet size of 8 Kbytes (i.e., 167 cells are transmitted for a single packet). All connections are with the same scheduling priority, thus the traffic being served in the switches is on a first-in-first-out basis. A positive acknowledgment for a packet

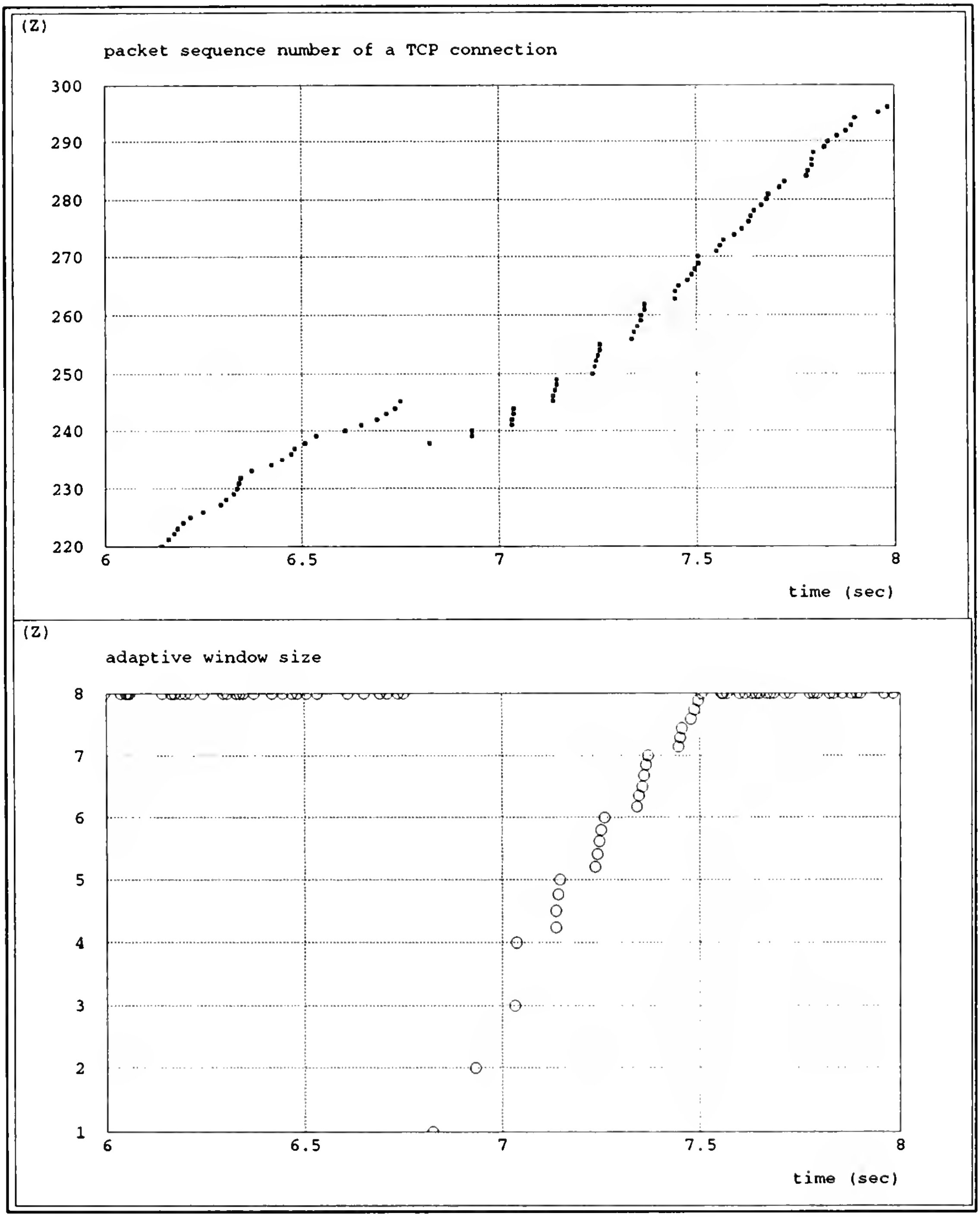


Figure 6.3: Traces of the packet sequence number and corresponding window size of a TCP connection

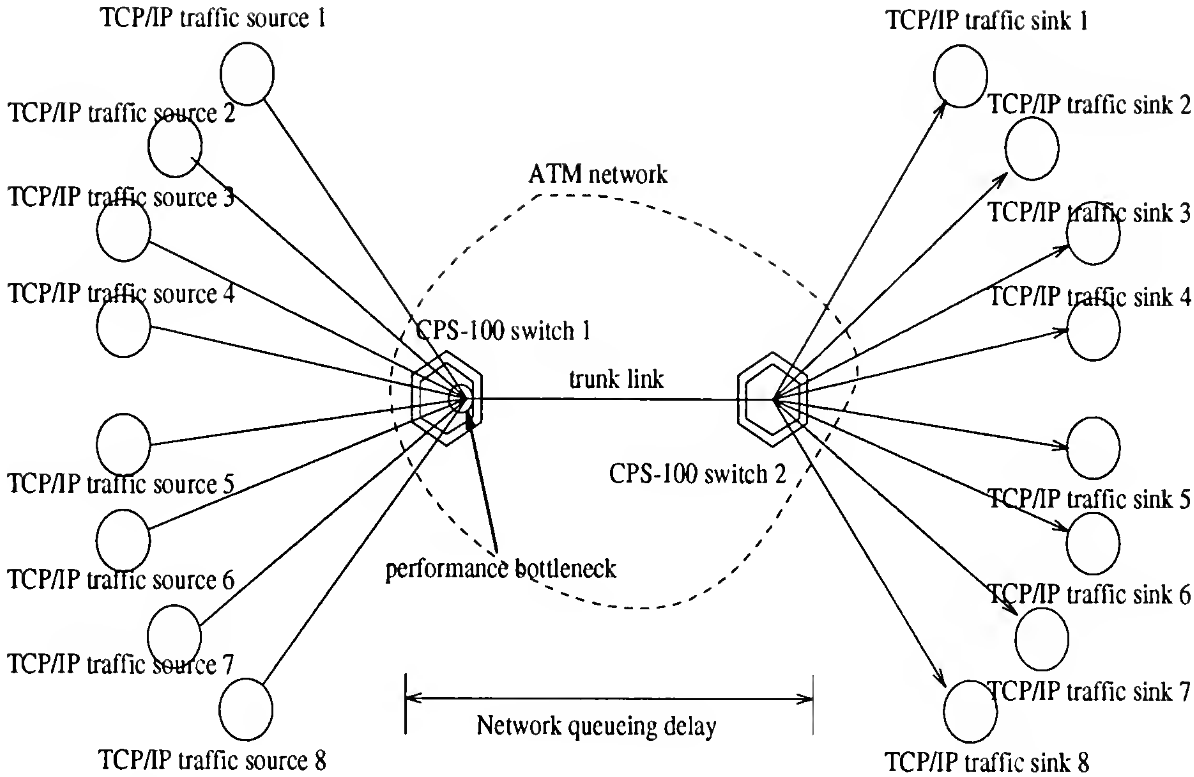


Figure 6.4: Network configuration: simulation scenario A

will be sent back to the sender after the receiver gets all the cells belonging to the packet in correct order. The round trip time (RTT) is composed of the round-trip propagation delay on the path, the round-trip network queueing delay and a processing delay of $100ms$ at the receiver. To simplify the simulated situation, the RTT is calculated by directly reading the measured average network queueing delay, so that premature retransmission due to an incorrect estimation of the RTT can be avoided. It is assumed that all packet losses are detected by the RTO timeout.

Suppose these TCP connections are initialized at some random time and, after an initial phase, all the connections have reached its maximum window size. The worst case that could happen is that within a short period all the connections decide to transmit long files to the receivers. Each traffic source tries to grab the whole link bandwidth when it transmits. The large burstiness of the incoming traffic causes congestion on the outgoing trunk port of switch 1 and make cell losses occur. Figure

6.5 presents the simulation results of the case. The traces of the packet sequence number of each connection are shown. In this simulation, all the link speeds are set to T3 (45 Mbps) and the round-trip propagation delay is assumed to be 200 cell times (about 1.88 msec). The egress buffer size at the trunk port is 500 cells. Five seconds of network time is simulated.

With a small amount of random delay at the beginning, each connection starts to transmit a full window of packets. As shown in Figure 6.5, most of the packets are corrupted by cell losses and these connections begin to retransmit packets. Since all the connections have the same scheduling priority, they will measure similar RTO intervals, which makes their retransmission process synchronize and results in a very high retransmission rate. It is observed that traffic synchronization phenomenon causes connections #1, #2, #4, #6, #7 and #8 retransmitting the same packet repeatedly and wastes a considerable amount of network resources. Note that the transmission could be very unfair. While some connections are waiting for the RTO timer to expire, some connections (e.g., connection #3 and #5 in this simulation) may luckily grab the available bandwidth and receive a high throughput. Figure 6.6 depicts the running connection goodputs that measure at the receivers. The connection goodput is updated whenever a receiver has obtained a good packet. It is seen that connection #3 and #5 achieve a connection goodput of 0.11 at the end of the simulation while the other connections are completely shut off.

6.3 Credit-Based Link-by-Link Flow Control: The N23 Scheme

The N23 scheme proposed in [68] is a method to implement credit-based link-by-link flow control. This scheme is performed independently on every virtual connection going through a particular link. An introduction of the N23 scheme is presented as follows. Further details of this control scheme can be found in [68]. At the beginning of transmission, the sender of the link is initialized with a credit balance of

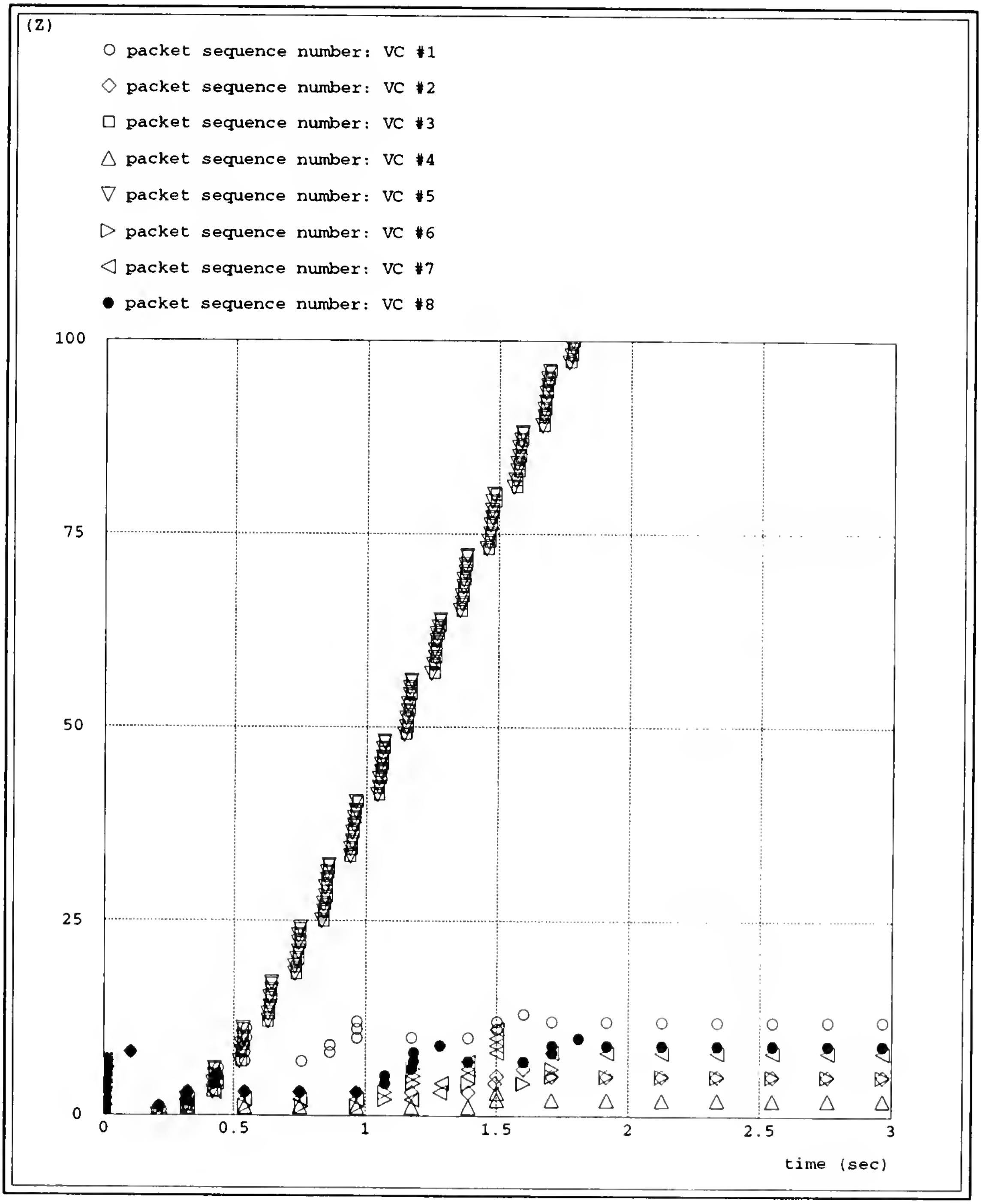


Figure 6.5: Traces of the packet sequence numbers of the TCP connections

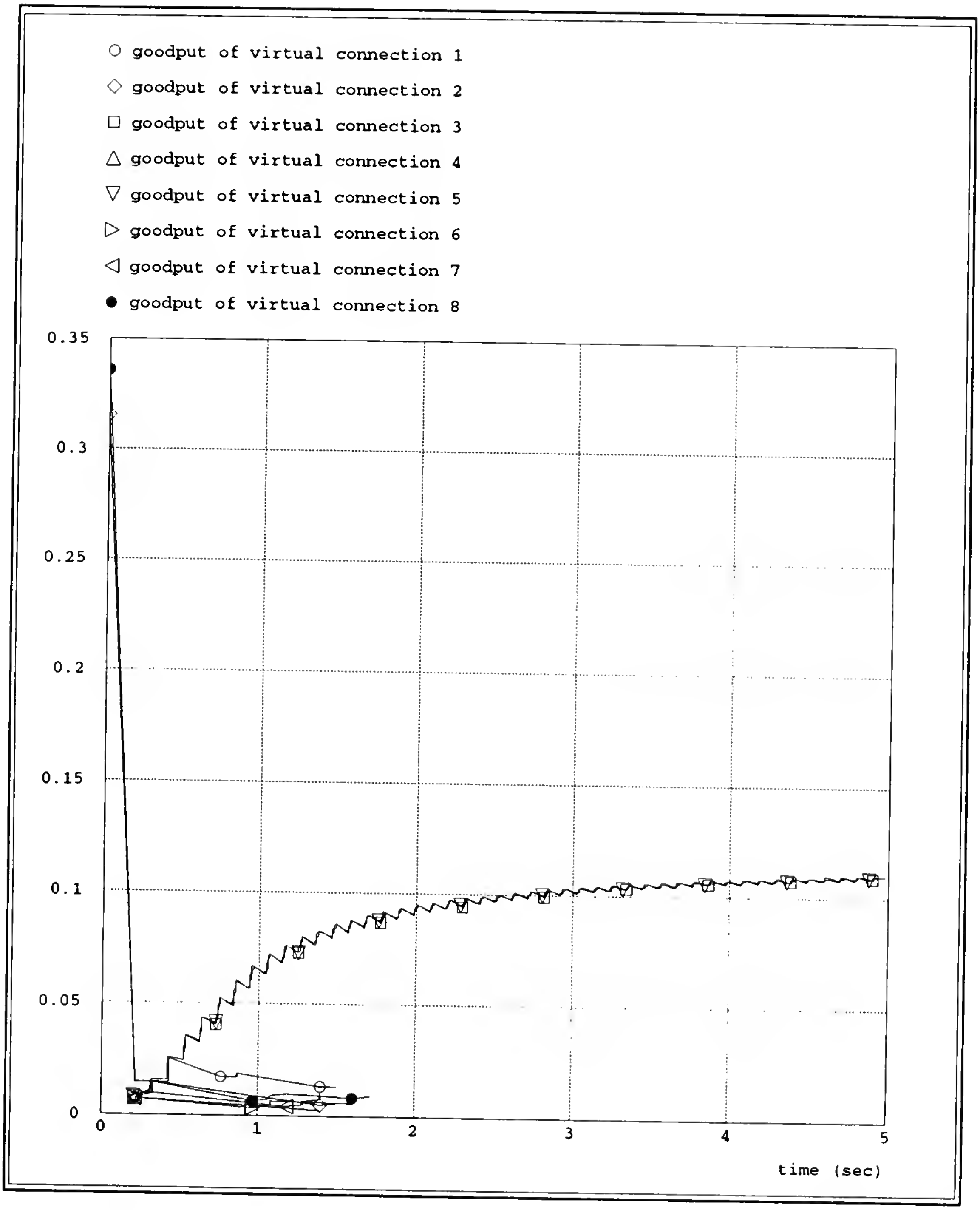


Figure 6.6: Running connection goodputs measured at the receivers: Simulation Scenario A

the available cell buffer for a connection at the receiver. The credit balance is decremented by one each time a cell of the connection is forwarded to the receiver. As the credit balance of the connection reaches zero, the sender has to stop sending cells. A credit cell with the updated information of the available cell buffer is returned to the sender every time the receiver has removed a certain number (= N2) of cells from the VC buffer. The sender updates its credit balance for the connection as it receives credit cells.

The available cell buffer of a connection at the receiver is divided into two zones, each having N2 and N3 cells respectively. The parameter N2 determines how frequently a credit cell should be returned to the sender. Current proposals suggest a N2 value of 10, which means 1/10 of the connection bandwidth in the reverse direction has to be used for transmitting credit cells. However, the bandwidth overhead for credit cells can be reduced by a factor of 6 if the credit information of different connections can be packed in a single credit cell. The size of the N3 zone is determined by the connection's bandwidth requirement and should be proportional to the round trip link delay (denoted by RTT_{link}) and the targeted average link bandwidth of the connection (denoted by B_{vc}), that is, $N3 = B_{vc} \cdot RTT_{link}$.

The N23 scheme can be implemented by the following procedures.

- Initiation:

credit_balance	=	N2 + N3;
V_s (cell count at the sender)	=	0;
V_r (cell count at the receiver)	=	0;

- Receiver side procedure:

  - When the receiver sends a cell of the connection to its outgoing link,

    V_r += 1;

  - For every N2 cells that have been sent, send a credit cell back to the sender with value V_r.

- Sender side procedure:
  - When the sender sends a cell of the connection to the receiver (if $credit\_balance >$ 0),

    $$\begin{array}{lll} V_s & += & 1; \\ credit\_balance & -= & 1; \end{array}$$

  - When a credit cell is received,

    $$credit\_balance \;=\; N2 + N3 - (V_s - V_r);$$

Since the sender is eligible to forward cells to the receiver only if it has credits, the cell buffer at the receiver will never be overflowed by the incoming traffic and thus zero cell loss can be guaranteed. The N23 flow control scheme is simulated on the single-congested-link network configuration shown in Figure 6.4, with the same network parameters used in Section 6.2. To examine the impact of propagation delay, it is assumed that the network propagation delay is dominated by the propagation time from the traffic sources to the congested point (i.e., the outgoing trunk port at the first switch). The egress buffer at the trunk port is allocated equally to each TCP connection and N2 is set to 10. The VC buffer of each TCP connection at the trunk port is served in a round robin fashion.

Figure 6.7 shows the steady-state credit variability at one of the TCP traffic sources. Note that in this simulation, the cell transmission is limited not only by the availability of credits but also the TCP adaptive window size. In general, the maximum window size of a TCP connection has to be proportional to the product of the targeted link bandwidth and the end-to-end RTT to achieve the targeted throughput. The sum of the TCP window sizes (= 64 Kbytes $\times$ 8 connections) in the simulation is slightly less than the product of the full link bandwidth and the end-to-end RTT [1]. Thus the credit balance is accumulated while the TCP connection

[1]The largest window size that can be used for a TCP connection is 65 Kbytes according to the previous TCP protocol. The later extensions have included a new TCP option to allow a larger window size [84].

is waiting for a packet acknowledgment to return. Once an acknowledgment has been received, the traffic source begins to transmit the next packet and consumes all its credits as soon as they are available.

Figure 6.8 depicts the running goodputs of each TCP connection. It is seen that in the steady state the connections equally share the maximum achievable bandwidth, which is slightly lower than the full link bandwidth due to the limit of the TCP window. There is no cell loss or TCP packet retransmission for any of the connections.

6.4 Rate-Based Flow Control: The BECN Scheme

The backward explicit congestion notification (BECN) mechanism proposed in [83] is a rate-based control scheme that uses feedback signals to adjust the cell transmission rate of each virtual connection. In the BECN scheme, the buffer utilization at network multiplexing nodes is monitored. When the buffer usage exceeds a threshold, further incoming cells from an active connection will trigger BECN cells to be transmitted from the congested node back to the active source. Upon receipt of a BECN cell, the source will reduce its current transmission rate by 50%. Thus if successive BECN cells have been received, the traffic source will reduce its transmission rate to 50%, 25%, 12.5%, etc., until a pre-defined lowest level of transmission rate. When the lowest transmission rate has been reached, the receipt of a further BECN cell will force the source to stop transmitting. The transmission rate will be recovered to its original value in a reversed way, specifically, by doubling the current transmission rate if no BECN cell is received during a predetermined source recovery period. To make certain of fairness between all sources, the source recovery period is chosen to be proportional to the current level of transmission rate. Thus when a traffic source is at a lower transmission rate, it will have a shorter recovery period.

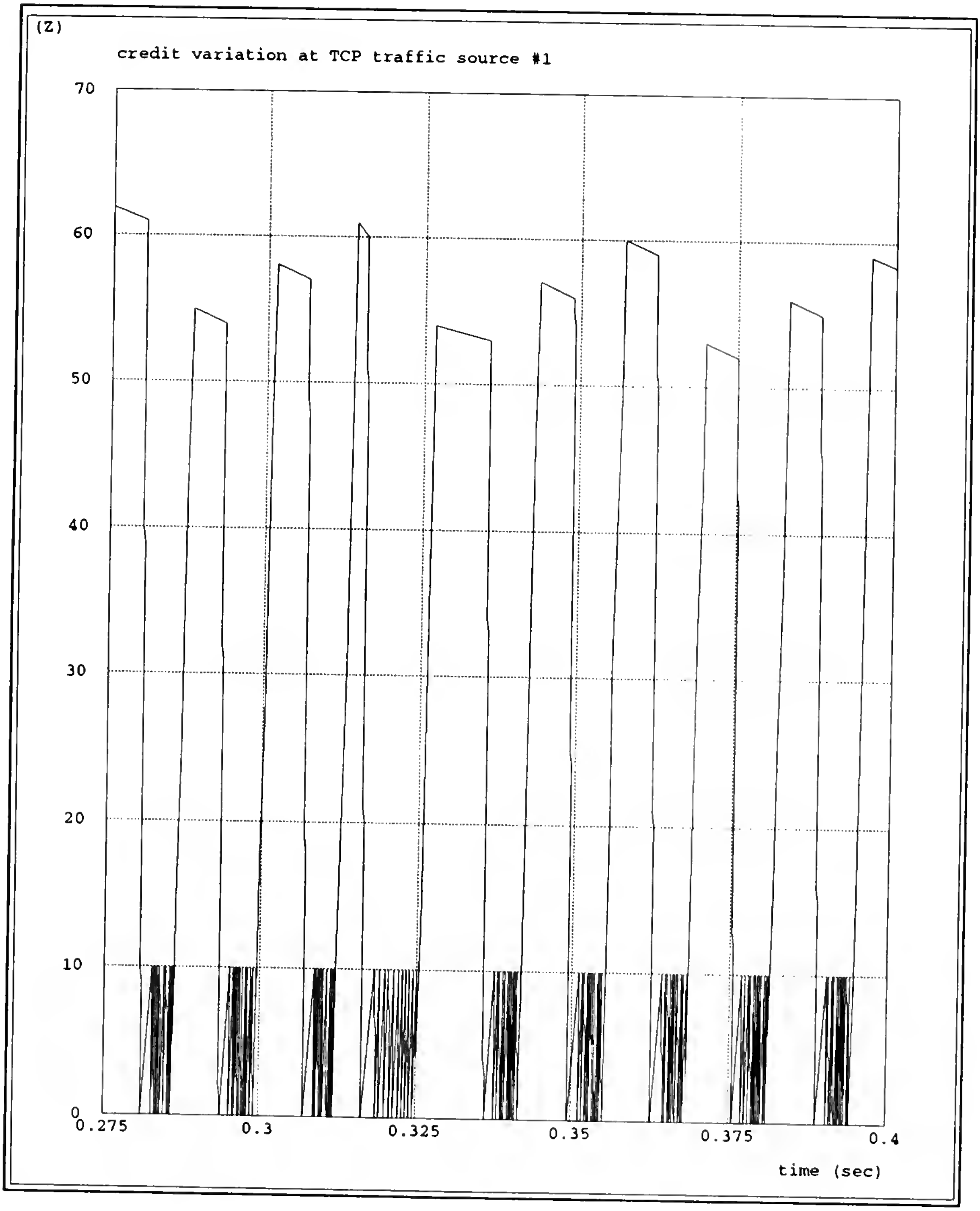


Figure 6.7: Steady-state credit variability at a TCP traffic source

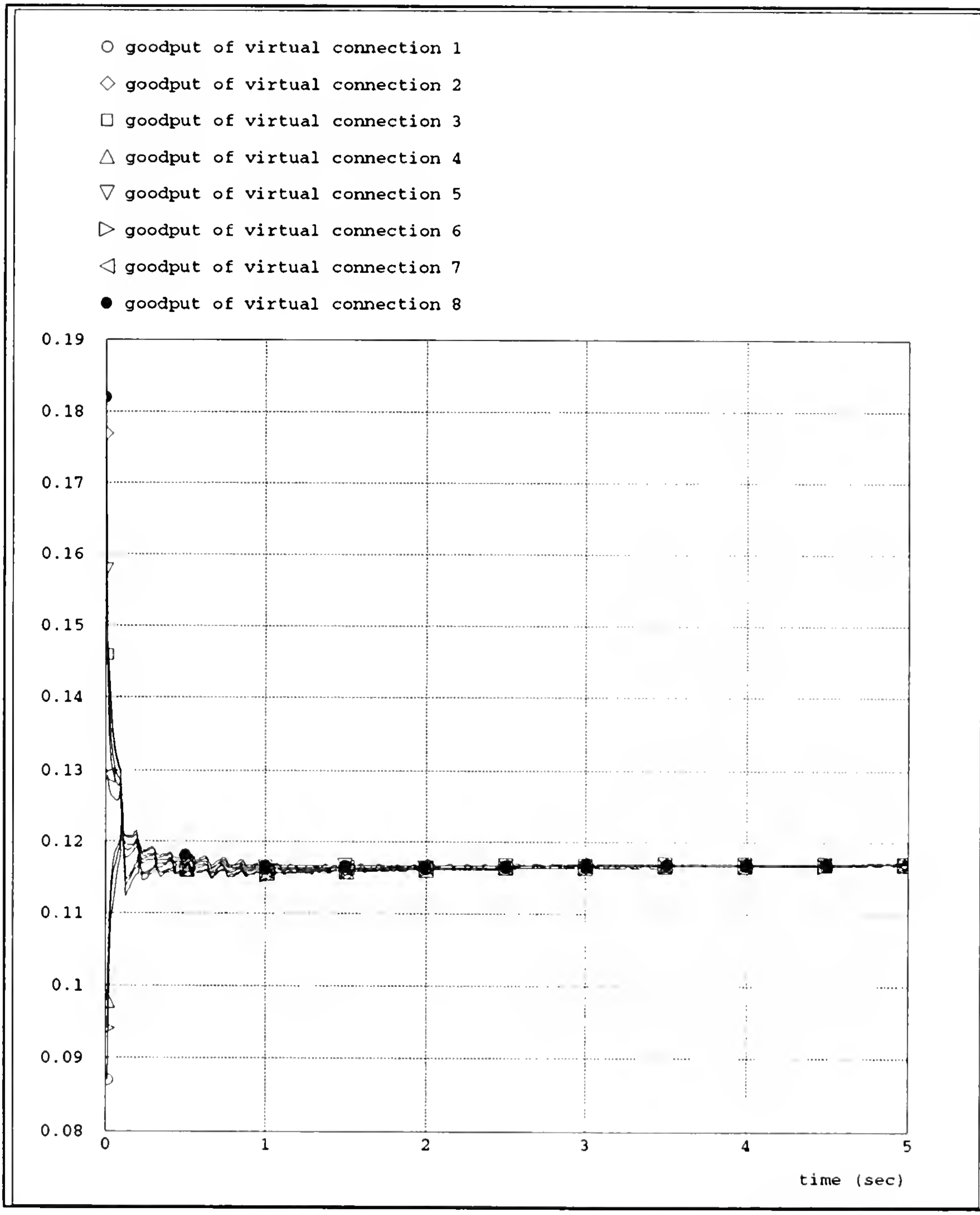


Figure 6.8: Connection goodputs with the credit-based control scheme: Simulation Scenario A

To reduce the transmission of excessive BECN cells, various filtering schemes have been proposed [83]. A per virtual channel filtering scheme, which is expected to produce the optimal system performance, is suggested. With this filtering scheme, as a buffer is congested, the switching node will transmit at most one single BECN cell back to an active source during each filtering time period. The filtering time period is set to be the same order of magnitude as the maximum propagation delay between the switching node and the traffic sources.

The BECN scheme is implemented by the following procedure:

- Initiation:
  set $throttling\_levels$,
  $buffer\_utilization\_threshold$,
  $minimum\_recovery\_time\_constant$.

  $$source\_recovery\_period = minimum\_recovery\_time\_constant \times 2^{throttling\_levels};$$

- Switching node procedure: when the buffer utilization exceeds the threshold, send a BECN cell back to the cell source upon the receipt of an incoming cell, if no BECN cell has been transmitted to the cell source during the filtering time period.

- Traffic source procedure:

  - When a BECN cell is received,
    $$\begin{aligned} source\_transmission\_rate &= source\_transmission\_rate \times 0.5; \\ source\_recovery\_period &= source\_recovery\_period \times 0.5; \end{aligned}$$
  - If no BECN cell is received during the current source recovery period and the transmission rate level is less than 100%,
    $$\begin{aligned} source\_transmission\_rate &= source\_transmission\_rate \times 2; \\ source\_recovery\_period &= source\_recovery\_period \times 2; \end{aligned}$$

With the control scheme, the active traffic sources will be forced to center at a transmission rate which is about its fair share of the available bandwidth in the steady state. Figure 6.9 depicts the peak cell rate variation of a traffic source

by simulating the BECN scheme on the single-congested-link network configuration, shown in Figure 6.4. It is seen that the source transmission rate fluctuates around a level of its fair sharing link bandwidth (i.e., 12.5% in this case).

As described in the above procedure, there are a number of system parameters that have to be set by the network for each connection. The effects of the different system parameters on the network performance are investigated in this section. The simulation is conducted on the single-congested-link network configuration shown in Figure 6.4, with the same network parameters described in Section 6.2, unless specified otherwise. For all the simulations, the number of the nonzero transmission rate levels for the traffic sources is assumed to be 8, including the 100%. Therefore, the minimum nonzero rate level is 1/128 (= 0.78%).

6.4.1 The Effect of Buffer Threshold

The effect of the egress buffer threshold at the congested node is studied. The system packet retransmission rate and aggregated goodput are analyzed with various buffer thresholds (see Figure 6.10). Two different egress buffer sizes, 500 cells and 2000 cells, at the outgoing trunk port are simulated. The recovery time constant at the lowest transmission level and the BECN filtering period are set to 700 cell times (= 7 $\times$ the one-way propagation delay).

It is observed that the system performance is sensitive to the threshold for a small egress buffer size. As the buffer threshold is increased, the traffic sources are allowed to transmit traffic in a more aggressive manner and thus the system throughput as well as the buffer utilization is likely to increase. In addition, a higher threshold results in a lower transmission rate of BECN cells. The link bandwidth consumed by the transmission of BECN cells is decreased from 10^{-4} to 10^{-5} in the simulation as the threshold is increased. However, a higher threshold also increases the probability

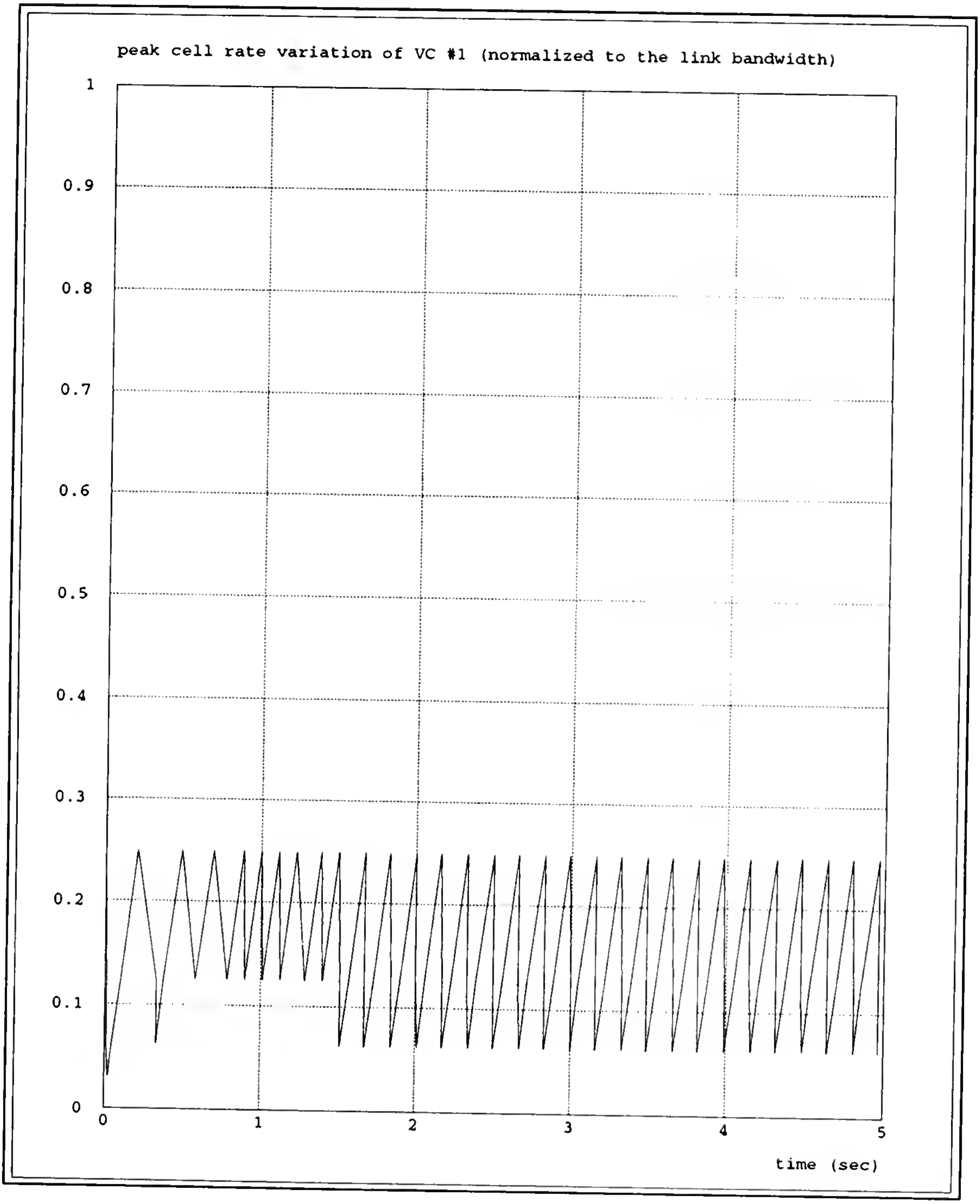


Figure 6.9: Peak cell rate variation of a traffic source with the rate-based control

of cell loss. As the threshold gets closer to the full buffer size, the packet retransmission rate becomes higher, inducing a lower system goodput. In general, a threshold around 50% of the buffer size will give a better network performance.

6.4.2 The Effect of Source Recovery Period

The minimum recovery time constant also has some effect on the system performance. With a buffer threshold of 50% and a one-way propagation delay of 100 cell times, different source recovery time constants are simulated on the same network configuration. The simulation result is presented in Figure 6.11. It is seen that for a large egress buffer, the variation of the recovery time constant doesn't show any significant impact on the system packet retransmission rate and aggregated goodput. On the other hand, for a small buffer, a lower recovery time constant will cause a higher packet retransmission rate. That is because when the traffic sources recover its transmission rate level more aggressively, the probability of cell loss increases at a much higher rate for a small buffer.

6.4.3 The Effects of Buffer Size and Propagation Delay

The buffer size at the trunk port and the propagation delay between the congested node and the traffic sources have significant impact on the system performance. A network model with various buffer sizes and propagation delays are simulated and the results are presented in Figure 6.12. For all the following simulations, the minimum recovery time constant and the BECN filtering period are set to 7 $\times$ the one-way propagation delay and the buffer threshold is set to 50%.

As the propagation delay between the congested node and the traffic sources increases, the time for the BECN scheme to take effect also increases. If the round trip delay time is about the same or larger than the time for a traffic source to transmit an entire window of packets at its peak rate, the probability of a vast packet loss in the window will be high and as a consequence, the entire window might have to be

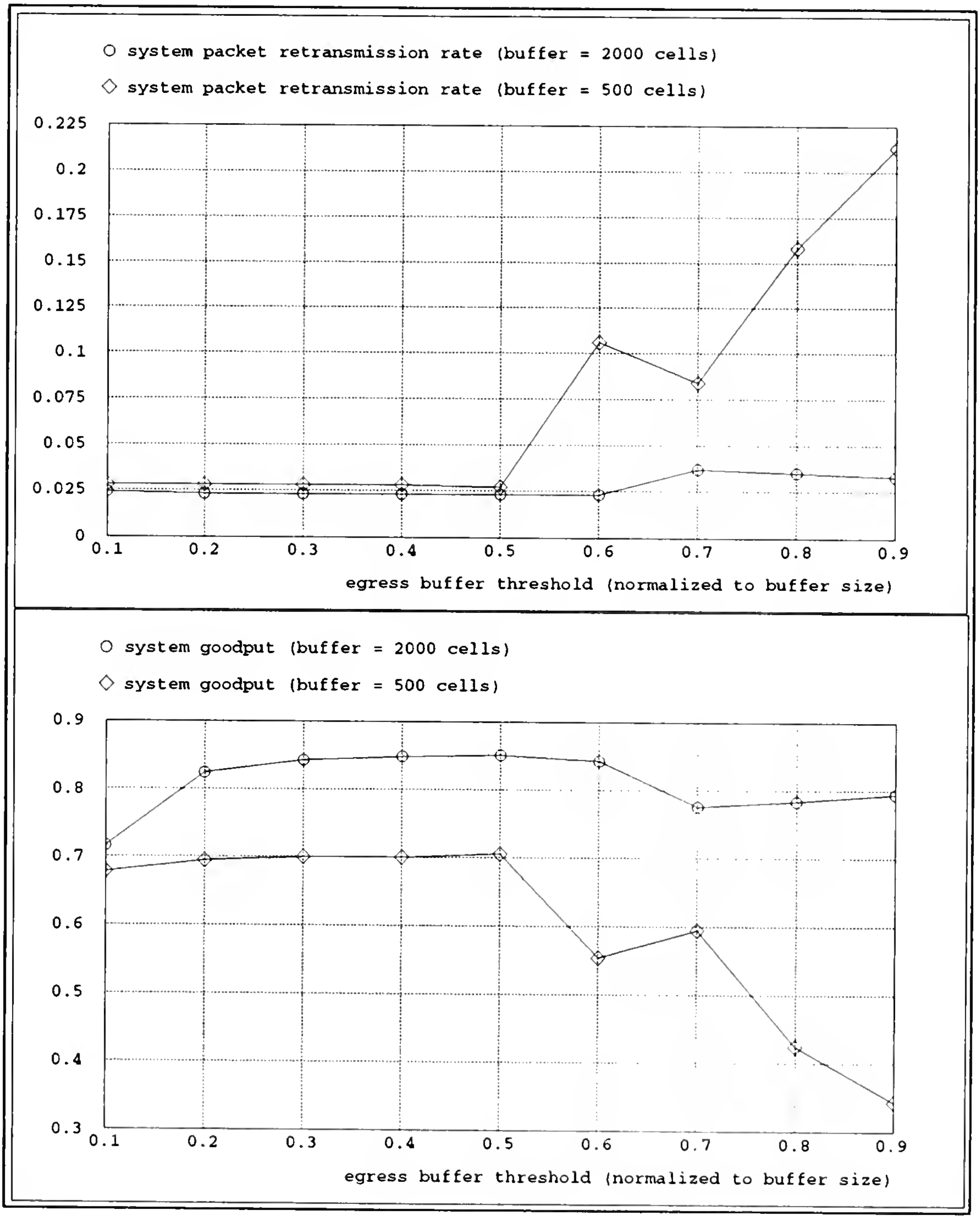


Figure 6.10: The system performance with various buffer thresholds

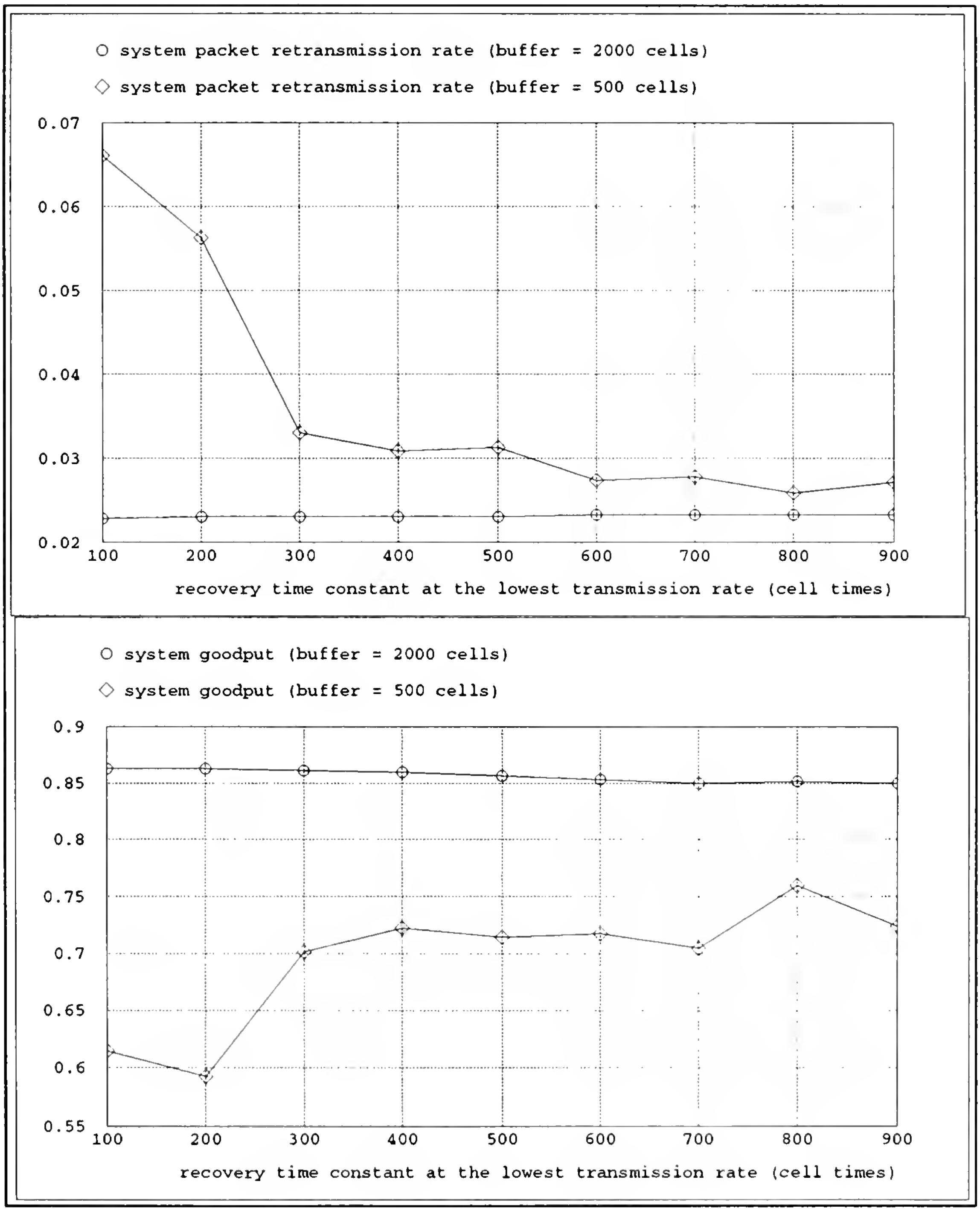


Figure 6.11: The system performance with different source recovery time constants

retransmitted. Figure 6.13 shows the traces of the packet sequence numbers of the TCP connections with a buffer size of 500 cells and a round trip propagation delay of 200 cell times. It is seen that due to the slow feedback of BECN cells, all the connections have to retransmit the packets in the first TCP window.

The packet retransmission rate and system goodput can be improved by using a larger buffer. A larger buffer size not only reduces the packet retransmission rate but also increases the speed for the system to reach a steady state. It is observed from the simulation that the BECN scheme has the effect to randomize packet transmissions of multiple TCP connections. As packets of different TCP connections are transmitted in a more randomized time frame, the traffic burstiness decreases and the buffer occupancy varies modestly. In consequence, each traffic source is able to recover gradually to its peak cell rate. In the steady state, the bandwidth utilization of each connection is actually limited by the TCP flow window size instead of source transmission rate. This explains why the system throughput can still be improved by a larger buffer size after zero packet retransmission rate has been reached.

The packet loss in the first window can be avoided by implementing a conservative *slow-start* procedure with the BECN scheme. Specifically, instead of allowing traffic sources to transmit at its peak rate from the beginning, the source transmission rate is forced to start with the lowest rate level and then recover to the maximum rate gradually. The same recovery procedure as in the BECN scheme is employed. With the slow-start, the traffic source increases its transmission rate conservatively to reduce the probability of potential congestion.

Figure 6.14 shows the simulation results of the slow-start BECN scheme. It is seen that the packet retransmission rate and system goodput are significantly improved for a limited buffer size. With a buffer size as small as 500 cells, zero cell loss and packet retransmission rate can be achieved regardless of the propagation delay. Note that the system goodput with a larger propagation delay is basically

limited by the TCP flow window size. The system goodput can be increased if larger TCP windows are employed. However, larger windows put higher demands on the network and thus increase the probability of potential network congestion.

6.5 Network Performance Comparison and Discussion

In this section, the network performance improvement with various flow control schemes is analyzed. The control schemes are simulated on two different network configurations and the connection goodputs are studied.

6.5.1 Simulation Scenario A

With the single-congested-link network configuration depicted in Figure 6.4 and the same network parameters that were specified in Section 6.2, the connection goodput of a TCP traffic source with different flow control schemes are compared, as shown in Figure 6.15. It is seen that the credit-based control provides the best network performance improvement among all the control schemes as expected. The connection goodput reaches a steady state after a short delay at the beginning. The connection goodput with the BECN scheme suffers a slow ramp up before achieving its steady state, mostly due to the slow-start TCP window control algorithm evoked upon packet loss. The slow-start BECN scheme improves the speed for the connection goodput to reach a steady state, which is about 0.025 of bandwidth less than with the credit-based scheme. The connection goodput can be improved further by making use of a larger buffer (e.g., 2000 cells). With a larger buffer, the slow-start BECN scheme is able to achieve the same steady-state goodput as with the credit-based scheme, after some delay at the beginning.

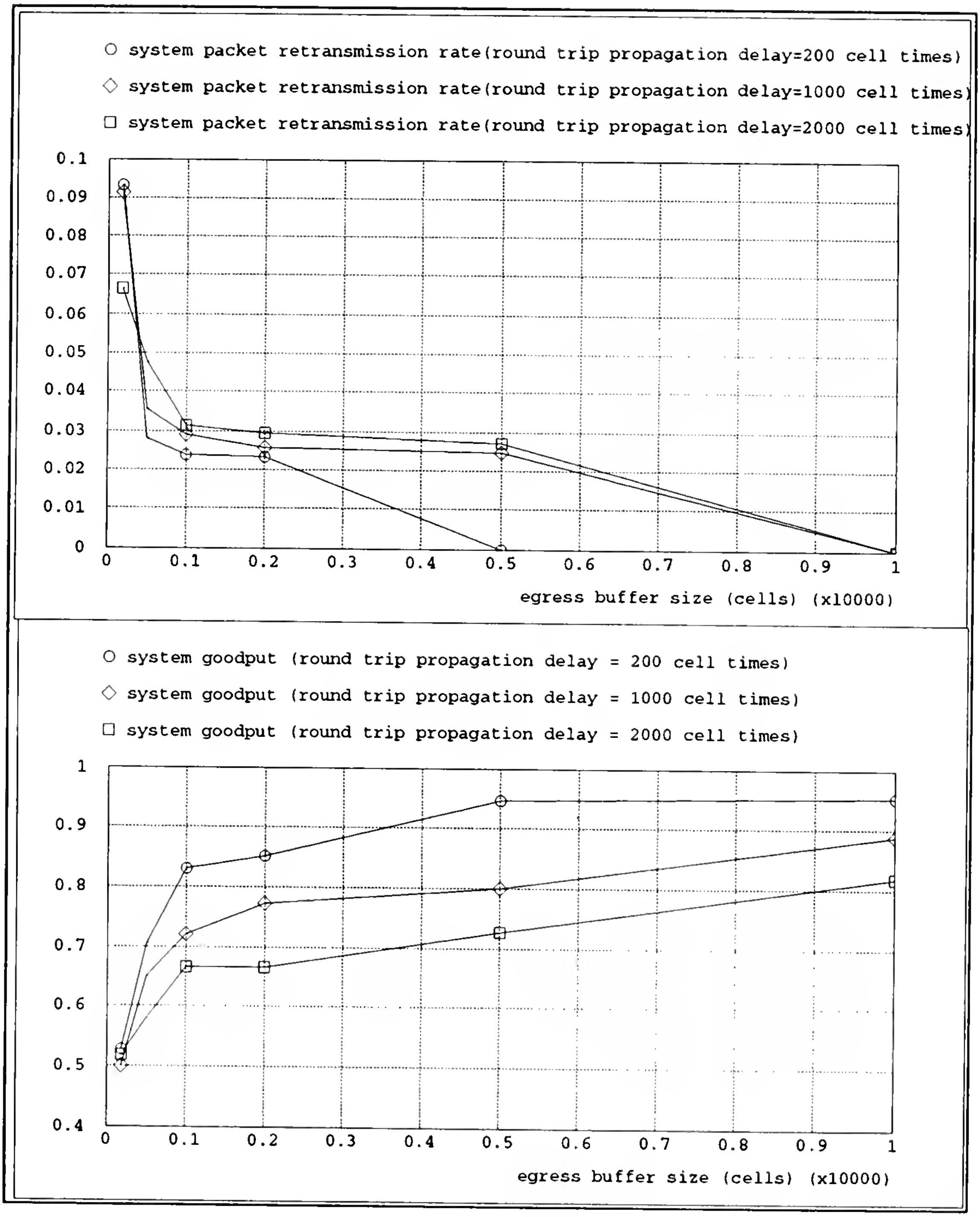


Figure 6.12: The system performance with different buffer sizes and propagation delay

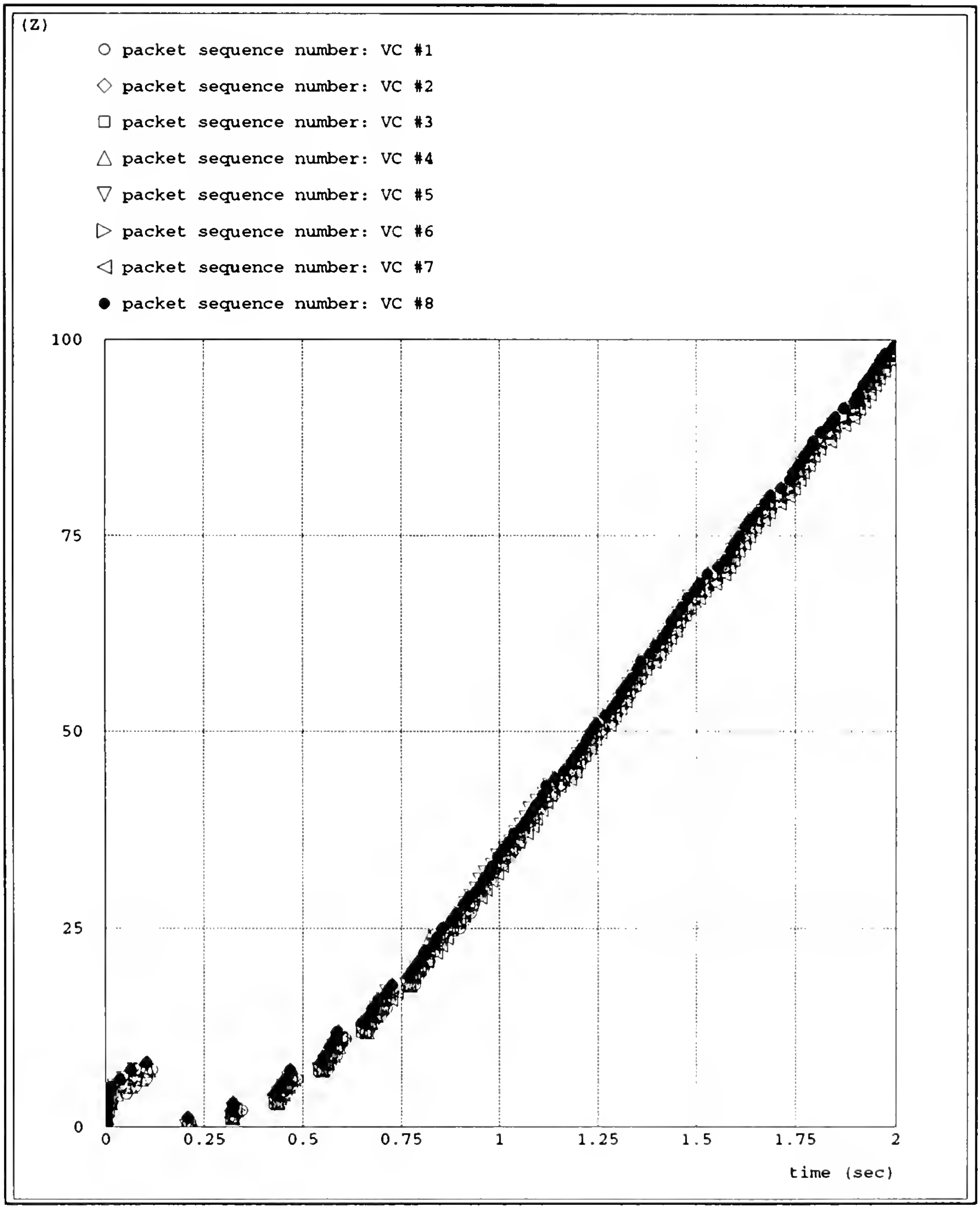


Figure 6.13: Traces of packet sequence numbers of TCP connections with the BECN scheme (buffer = 500 cells and round trip propagation delay = 200 cell times)

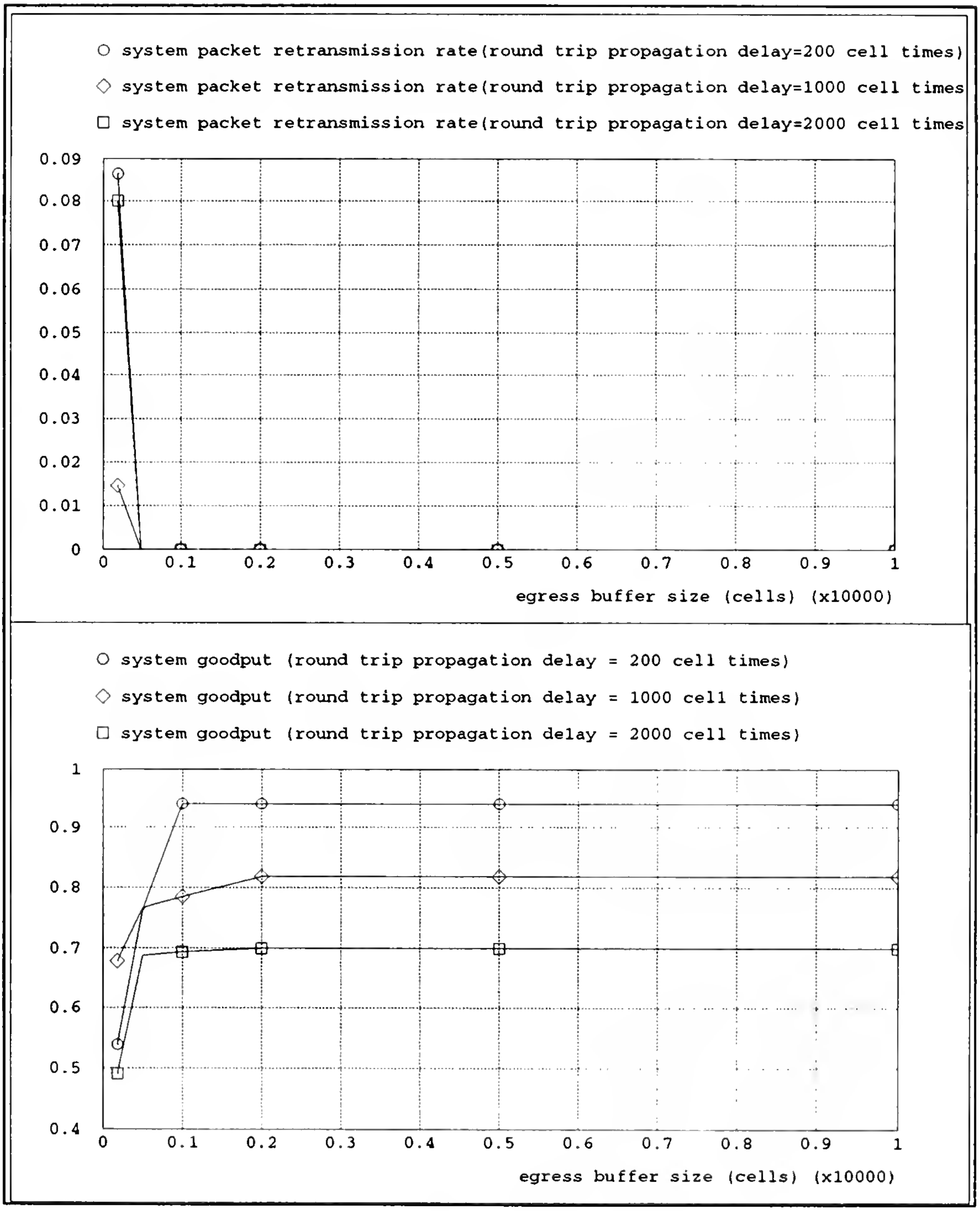


Figure 6.14: The system performance with the slow-start BECN scheme

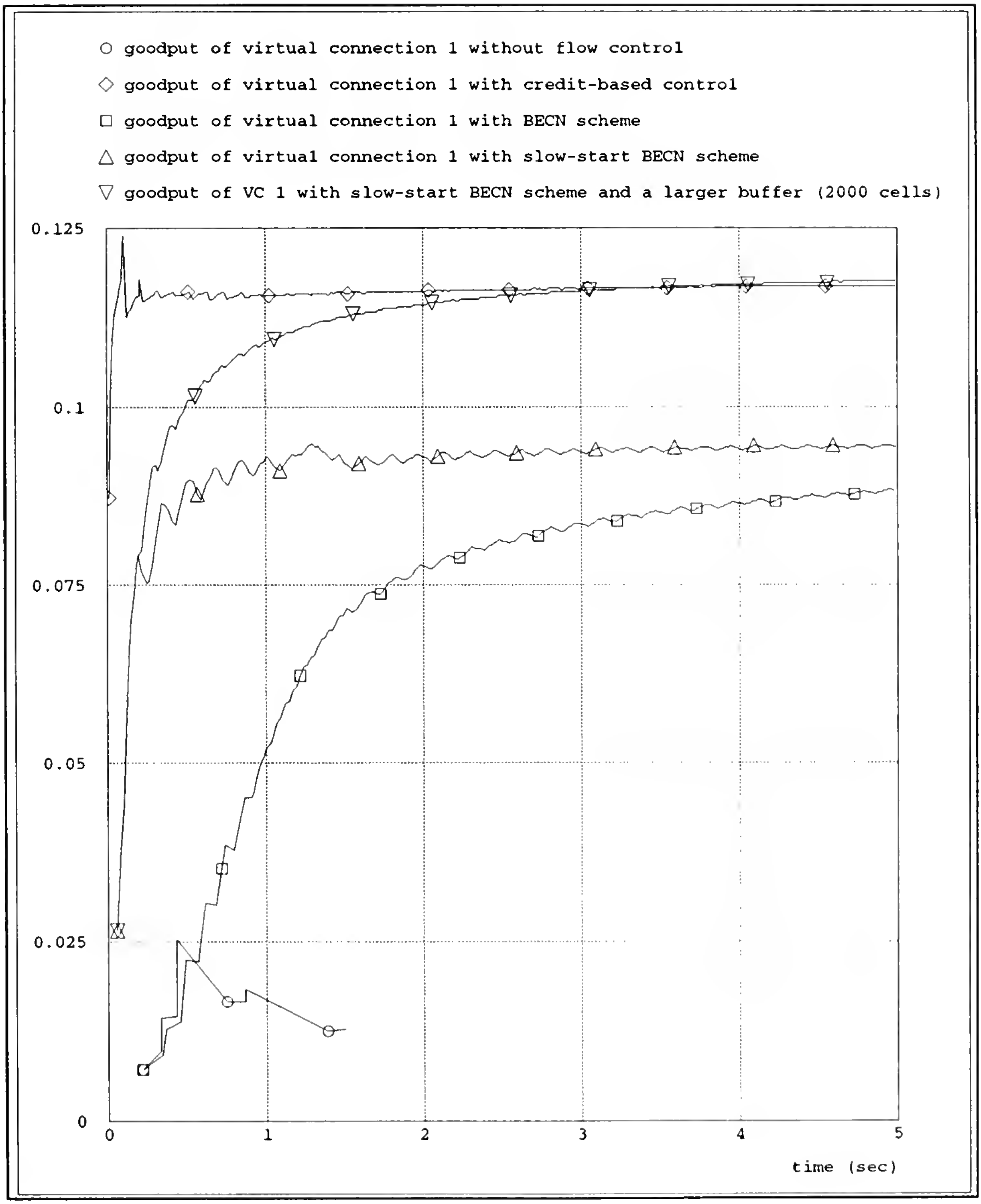


Figure 6.15: The connection goodput with different flow control schemes: Simulation Scenario A

6.5.2 Simulation Scenario B

In this section the network performance on a more complicated network topology is considered. This network configuration (see Figure 6.16) consists of two bottleneck links interconnecting three switches. There are totally four and six virtual connections competing for the outgoing trunk links of switch 1 and switch 2, respectively. Assuming all the links are using the same rate, trunk link 2 becomes a more congested link, where each connection is expected to receive a fair share (i.e., 1/6) of the available bandwidth. Since the connection bandwidth of virtual connection 1 and 2 is limited by trunk link 2, connections 3 and 4 are eligible to share the remaining link bandwidth of trunk link 1, which is 1/3 for each connection.

To utilize fully the maximum link bandwidth, the TCP maximum window size is increased to 512 Kbytes with a constant packet size 64 Kbytes (i.e., 1334 cells are transmitted for a single packet). It is assumed that the round trip propagation delay between the traffic sources and the first congested node on their paths, and between the two congested nodes, is 200 cell times each. The egress buffer size at each trunk port is 500 cells.

Figure 6.17 shows the simulation results with the credit-based control. It is seen that the goodput of each the TCP connection converges quickly to its expected bandwidth utilization. Figure 6.18 presents the results with the BECN scheme. The TCP connections suffer performance degradations due to the slow feedback of the BECN cells. Most of the connections retransmit the packets in their first TCP windows. At the end of the simulation, connections 3 and 4 achieve a running goodput of about 0.22, while the connections competing for the second congested link receive a goodput of 0.11 on the average. The network performance and the fairness between different TCP connections can be improved with the slow-start BECN scheme, as shown in Figure 6.19. However, in this network configuration, connections 1 and 2 receive less bandwidth than the other connections going through the second congested

link (about 0.04 of bandwidth less). This is because they have two congested nodes on their paths and therefore they have a larger probability to receive BECN cells during transmission. A better performance can be achieved by using a larger buffer. Figure 6.20 shows the simulation results with a buffer size of 5000 cells. It is observed that at the end of the simulation, connections 3 and 4 achieve a running goodput of 0.28 while connections 1 and 2 have a goodput of 0.11 and all the other connections have a goodput of 0.16.

In addition, a link-by-link BECN scheme described in [58] is simulated on the same network configuration. In this control scheme, as the buffer in a switching node (e.g., switch 2) is congested, BECN cells are sent to its upstream nodes (e.g., switch 1) to throttle their transmission rate, thus creating backpressure on the traffic sources. The scheme avoids per-VC buffering and reacts to congestion on a link-by-link basis. However, the simulation results (see Figure 6.21) show the bandwidth utilization on trunk link 1 is degraded and connections 3 and 4 receive a much worse throughput than with the end-to-end BECN scheme. That is because the connections share the same output buffer at switch 1 with connections 1 and 2. Thus when the transmission rate of the outgoing port is throttled, the throughput of connections 3 and 4 are also affected. It has been proven that the link-by-link BECN control scheme may not perform well in a complex network configuration.

A fair comparison of the network performance improvement of credit-based and rate-based flow control has been performed. The simulation results clearly demonstrate that the control schemes are able to offer substantial performance improvement for TCP traffic over ATM networks. The impact of several system and network parameters for the rate-based scheme have been analyzed. By performing the simulation experiments on two different network configurations, the issues of the connection transient behavior and the fairness of resource sharing are investigated. It has been shown that the performance of the rate-based approach can be

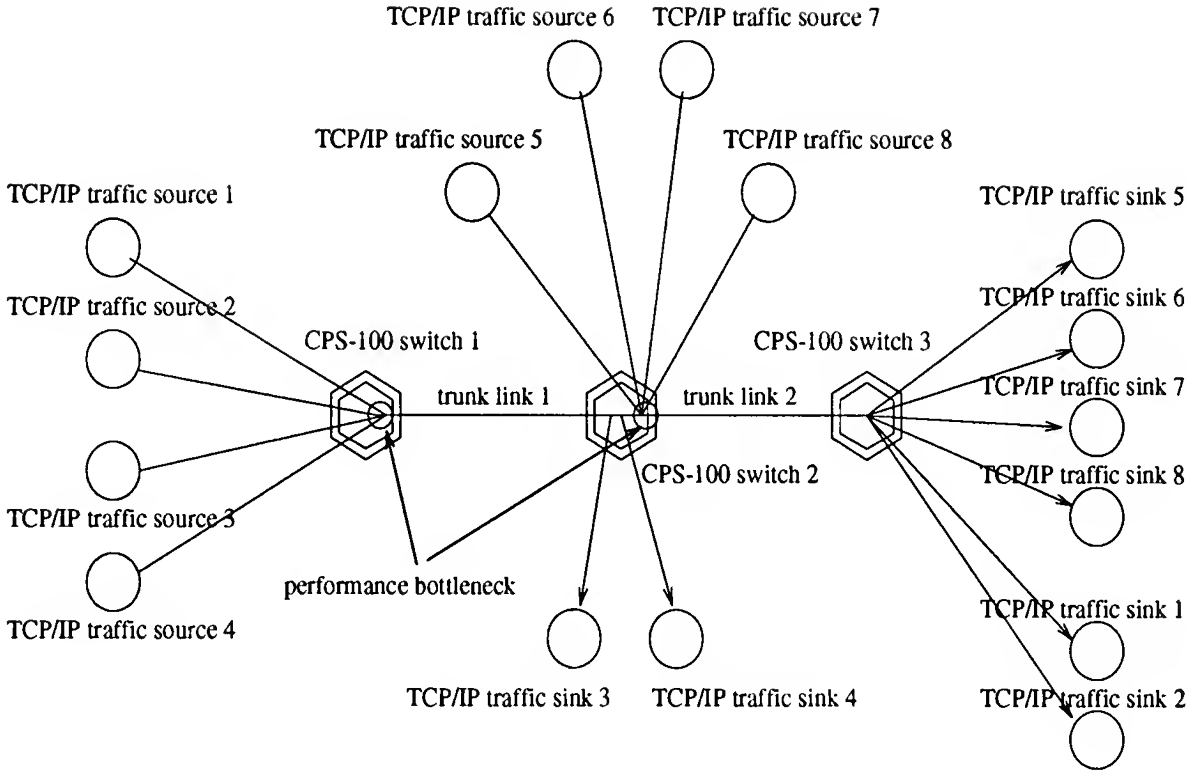


Figure 6.16: Network configuration: simulation scenario B

improved to a level comparable with the credit-based approach by making use of a slow-start procedure and a larger buffer. The study provides a beneficial solution for the switches that are not able to implement the significant amount of specialized hardware required by the credit-based approach.

It might be possible that the ATM Forum will select a single control approach while keeping the other one as a option. The network providers can then offer customized network services according to a users' QOS requirements. The credit-based scheme can be implemented to provide users with the per-virtual-connection traffic control and performance guarantee. For the users or applications that prefer an economical service and are also prepared to accept the network performance degradation under extreme variations of traffic load, it might be sufficient to employ a rate-based control scheme.

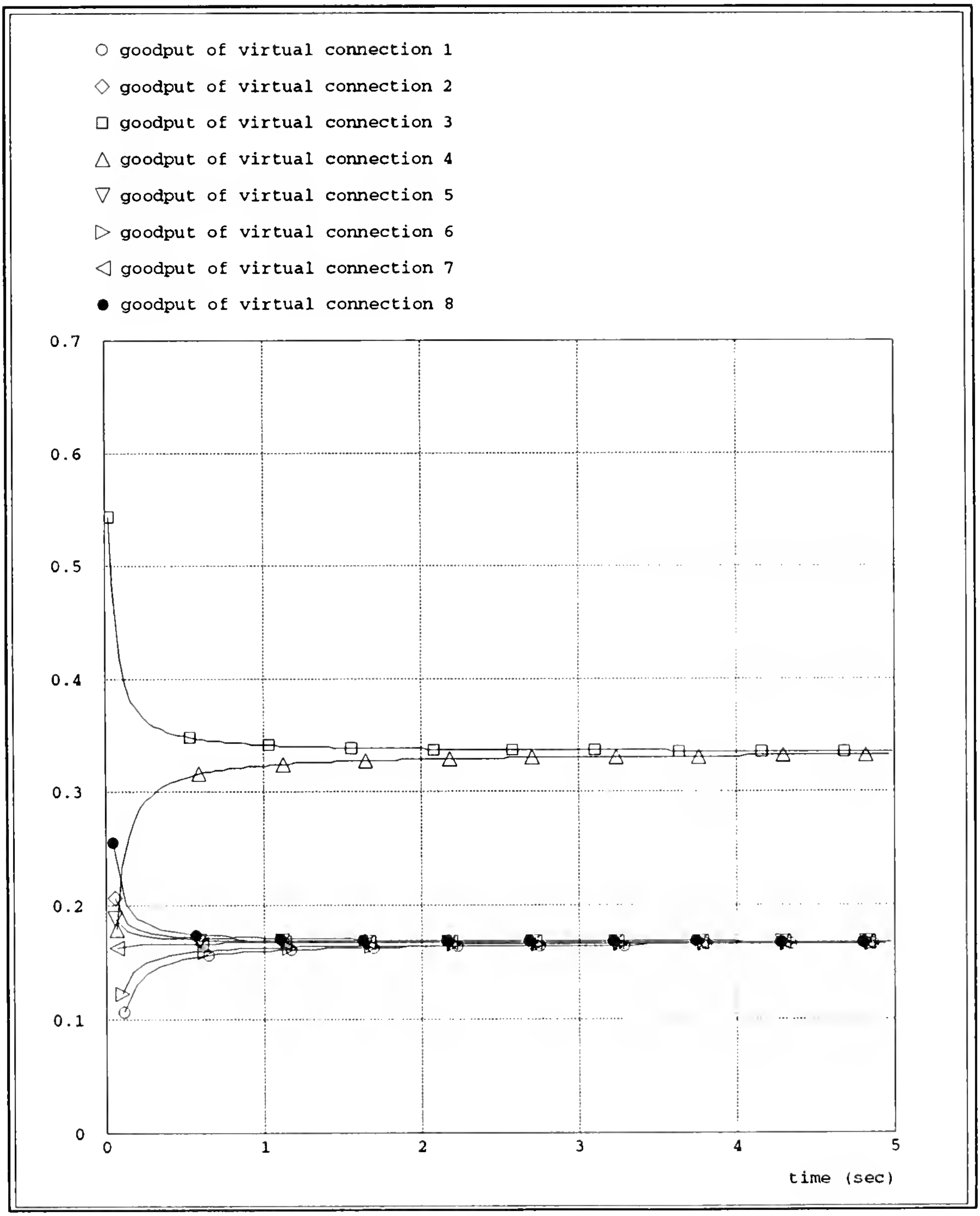


Figure 6.17: Connection goodputs with the credit-based flow control scheme: simulation scenario B

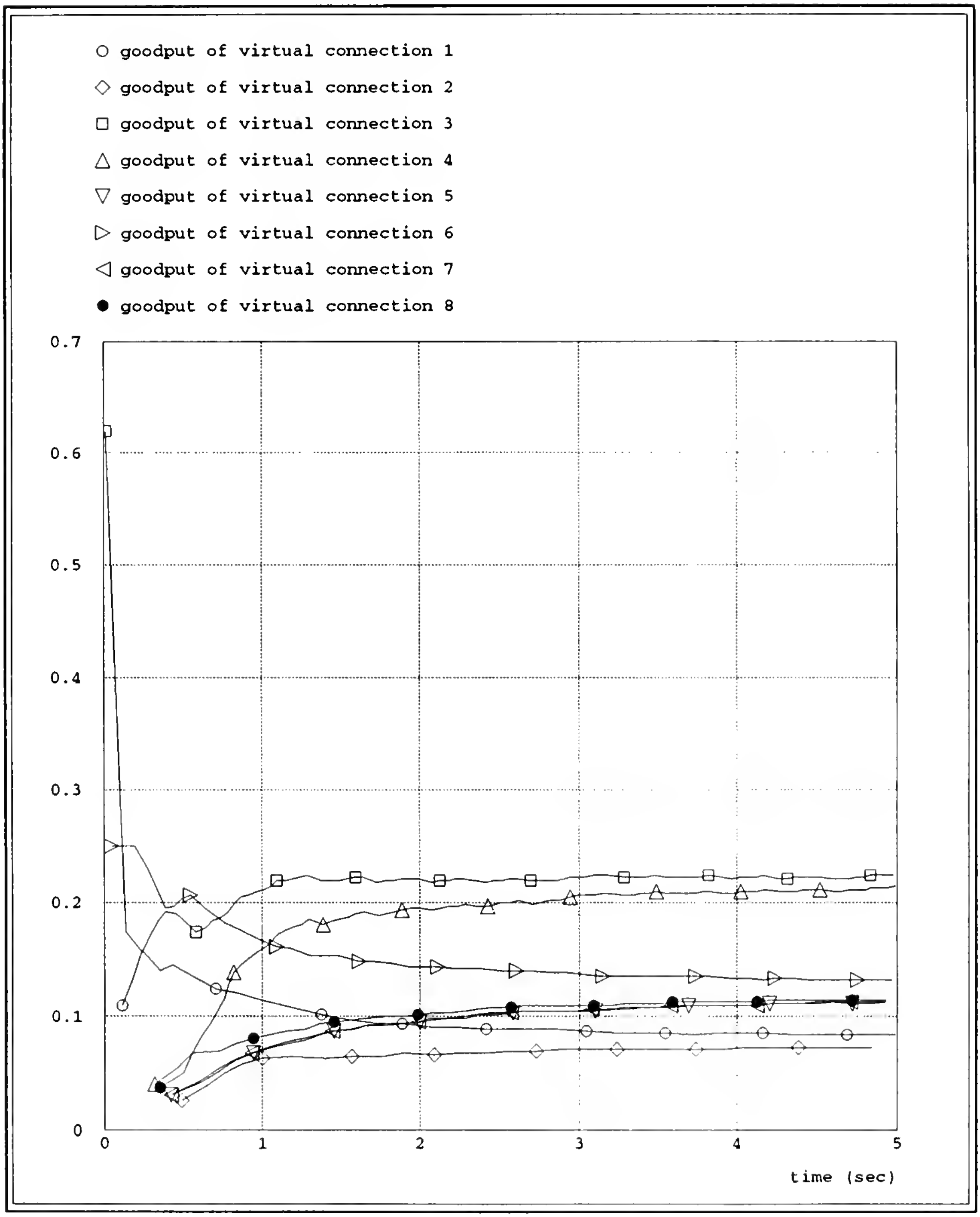


Figure 6.18: Connection goodputs with the BECN flow control scheme: simulation scenario B

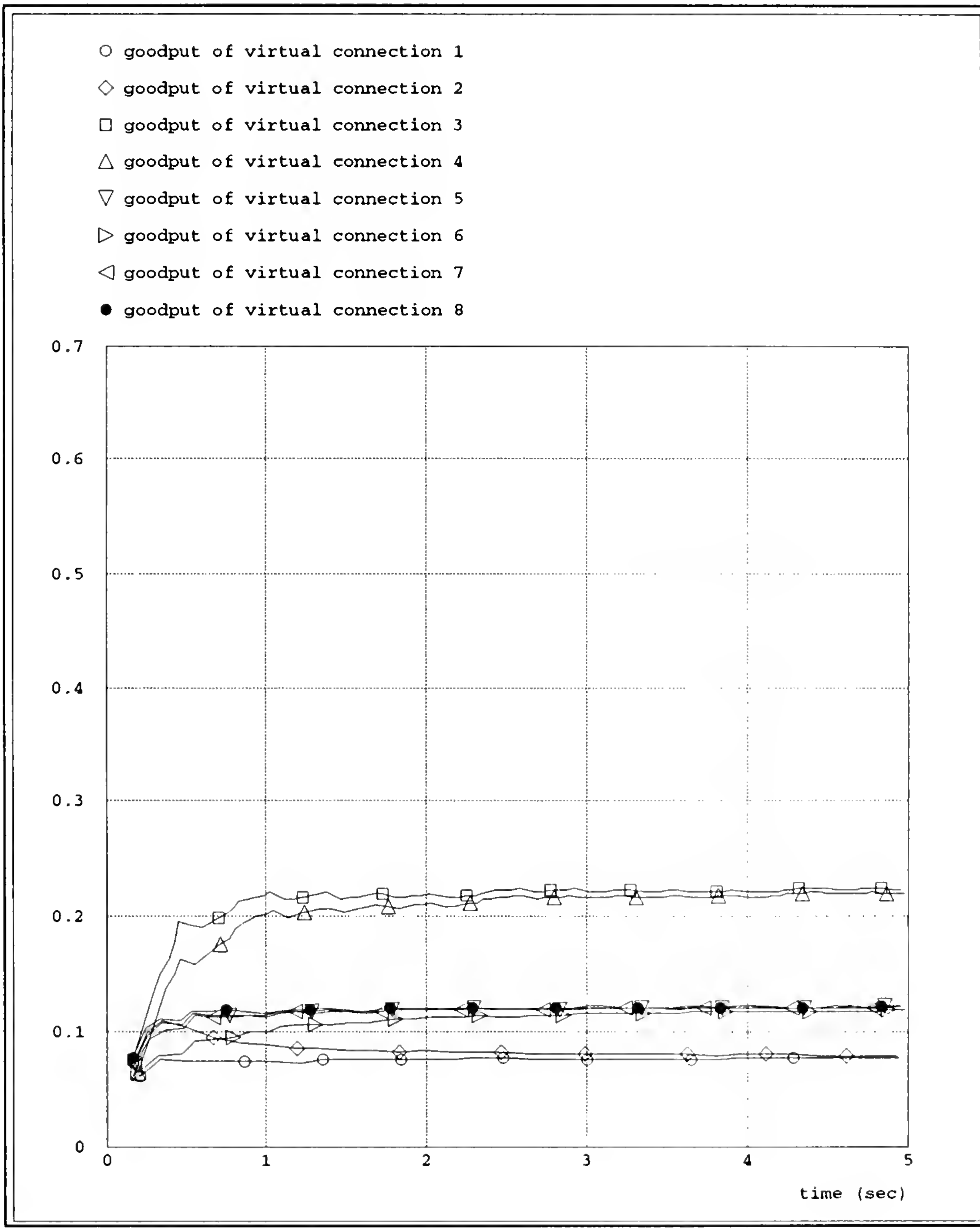


Figure 6.19: Connection goodputs with the slow-start BECN flow control scheme: simulation scenario B

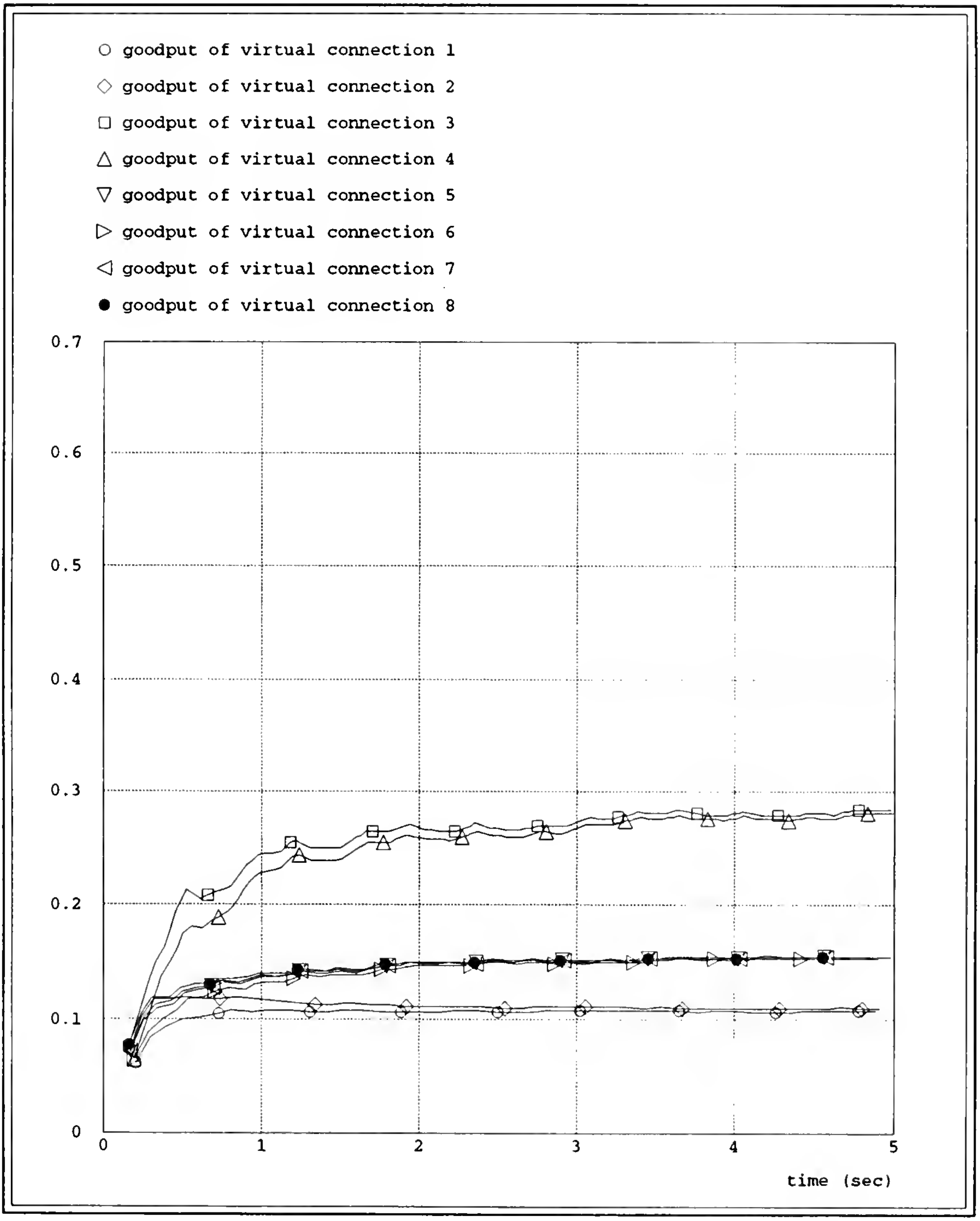


Figure 6.20: Connection goodputs with the slow-start BECN flow control scheme and a large buffer (5000 cells): simulation scenario B

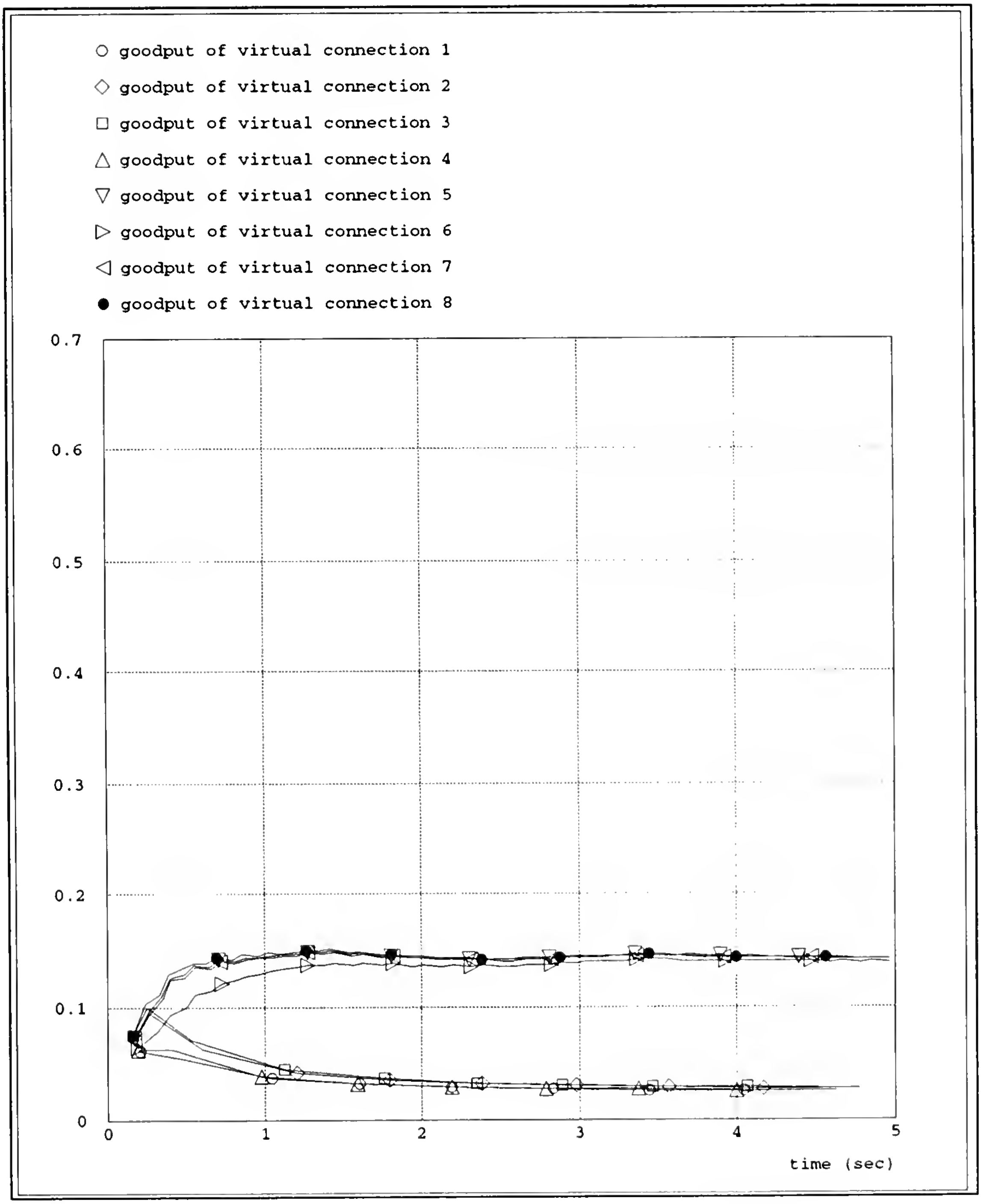


Figure 6.21: Connection goodputs with the link-by-link BECN scheme: simulation scenario B

CHAPTER 7
CONCLUSION

7.1 ATM Traffic Requirements

In this dissertation, we have considered the joint problem of ATM switch design and traffic control management for ATM networks consisting of such switches. ATM technology is being propelled by the need to provide a new generation of networking technology characterized by a single, unified and high-speed infrastructure capable of supporting all types of communication services. In the past different types of traffic and services have been supported on disparate networks and thus these networks could be optimized individually to meet their performance objectives. The drawbacks of this approach are that network providers need to build and maintain multiple networks and users must obtain different network interfaces to access the services. ATM technology offers a flexible means to multiplex different service types on a unified network and provides a significant efficiency improvement through the statistical resource sharing by multiple ATM connections.

However, the introduction of ATM also provokes critical issues on the design and implementation of network equipment, traffic management, signaling architectures, networking and end user protocols. For a successful deployment of the ATM technology in the broadband integrated services digital network (BISDN), it is necessary to design packet switches capable of switching relatively small packets at extremely high rates and supporting traffic management functions to provide adequate service guarantees for the large spectrum of anticipated applications.

The traditional reactive traffic control schemes, which typically depend on simple feedback signals such as negative acknowledgments from the receiver, do not

perform well in an ATM environment because of the large distances, high speeds and short cell transmission times involved. To support the diverse service requirements of users in a robust and flexible manner, it is essential for ATM networks to employ avoidance traffic control mechanisms so that an unacceptable level of network congestion can be prevented by means of conservative admission policies and resource allocations. The network congestion problem can be controlled in an effective and efficient way by making use of the congestion avoidance controls and employing reactive control schemes as backups.

7.2 Contributions of the Research

Many research efforts have been devoted to the development of traffic admission control schemes that manage traffic loading in the network by negotiating and allocating network resource to a connection in advance of transmission. Such control methods are well-suited to the class of traffic with predictable statistical characteristics. However, the issue of traffic controls for unpredictable and highly bursty traffic, such as data applications, and the issue of critical resource management and maintenance in a heterogeneous network environment have not been adequately treated. Additionally, although many studies have been conducted in the areas of ATM switch design and traffic management, the research in these two areas have progressed essentially independently. Few of the previous studies of ATM traffic controls have considered the feasibility and complexity of implementing a particular control mechanism on a realistic switch architecture. The research in this dissertation addresses these issues and bridges the gap between ATM switch design and traffic management.

Our considerations in this dissertation were motivated by a shared medium/output buffering ATM switch, designed by Loral Data Systems. As a beginning of the research, we conducted a complete study on the ATM switching technology and the

major approaches of ATM switch design. We identified the performance characteristics, limitations and potential application areas of three fundamental ATM switch architectures: shared medium, shared memory and space division. We developed a switch simulator, realizing accurately the hardware design of the Loral CPS-100 fast packet switch, as an experimental test-bed for our study on ATM switch design and traffic control. During the research period, we have performed numerous simulations and analyses to identify the switch characteristics with different traffic models and evaluate implementation complexity as well as effectiveness of various traffic control schemes. In this dissertation, we presented part of our simulation results to depict the basic switch characteristics, and demonstrated that the critical delay performance can be controlled effectively with the implementation of appropriate traffic control schemes at the network access and switching nodes.

We presented a complete analysis on the traffic characteristics, QOS requirements, modeling methods and potential ATM applications of three basic ATM service categories: CBR, VBR and ABR. Based on the analysis, we proposed an integrated hierarchical traffic control framework for the fundamental classes of ATM service. The traffic control architecture allows cooperative control actions at distributed network systems and provides a robust and complete control operating on different time scales.

Cell loss probability is another important performance criterion that is encountered in satisfying the needs of ATM services. However, very little work has appeared in the literature on the control of cell loss probabilities for different priority classes. We conducted a thorough investigation for a space priority control scheme to manage optimally a finite shared buffer system. The study produced a feasible and attractive means to support an arbitrary number of loss priority classes in an ATM switch. We developed a queueing model to characterize analytically the system, and present efficient optimization procedures that are capable of finding the optimal loss

thresholds to maximize system admissible load. We verified the optimization procedures, demonstrated the resource efficiency improvement and evaluated the impact of given traffic conditions and cell loss criteria by numerous numerical examples.

In addition, we considered the problem of traffic controls for ATM best-effort service. We identified the critical issues that are encountered as transmitting ABR traffic such as TCP/IP applications over ATM networks. An exhaustive comparative study has been performed on two flow control schemes that were proposed to the ATM Forum: the credit-based and rate-based schemes. We showed that with the implementation of the flow control schemes, packet retransmissions can be greatly reduced or completely eliminated, thus providing a substantial improvement on the network throughput. Because of the complexity in practical implementation of the credit-based scheme, the ATM Forum has decided recently to support the rate-based approach. We conducted a complete analysis on the impact of the system parameters that need to be specified for the rate-based scheme. We studied the issues of connection transient behaviors, fairness of resource sharing, and the impact of complicated network configuration with these two control schemes. Moreover, we demonstrated that with the implementation of a slow-start procedure and a larger buffer, the performance of the rate-based scheme can be improved to attain asymptotically the same level of performance as the credit-based scheme. The study produced a beneficial solution for switches that are not prepared to implement the credit-based control scheme.

In conclusion, this research work offers higher resource efficiency, improved network performance, and better resource protection for the ATM networking technology. We proposed effective traffic control mechanisms to resolve the critical issues encountered in ATM networks and, in addition, by considering a realistic switch architecture, we ensured that the proposed control schemes are fully acceptable for realization in ATM switch systems and networks.

7.3 Future Research

Considering the rate at which new networking applications are emerging, there is much scope for further research in this area. The following directions are suggested as a continuation of this research.

Traffic modeling. As new communication services evolve, network traffic characteristics as well as users' demands are expected to change. Network systems must response by modifying system architectures or traffic control functions to satisfy the needs. The problem of network resource management can be solved more efficiently if the resource requirements and traffic behavior of users can be accurately predicted. Therefore a careful investigation on the statistical nature of new networking applications and the development of traffic models that are capable of reflecting the characteristics play important roles for the ATM networking technology to success continuously. For further simulation study of ATM traffic controls, accurate traffic models should be employed to synthesize realistic traffic and to yield precise network performance predictions.

An extension of the rate-based flow control. Recently Barnhart [85] proposed a Proportional Rate-Control Algorithm (PRCA) which provides a further refinement on the rate adjustment mechanism employed in the original BECN scheme. The scheme is currently under investigation by the ATM Forum. With the PRCA, the allowed cell transmission rate decreases every time a cell is transmitted by an amount proportional to the current cell rate. In steady state the allowed cell rate is restored to its previous balanced level as the traffic source receives a positive Resource Management (RM) cell, which is sent by the destination periodically as no congestion occurs. Note that using RM cell is only an alternative implementation of the FECN/BECN mechanism, by which a positive indicator is sent to notify there is no congestion.

While our study provides a close insight into the basic BECN scheme, the study can be extended to incorporate the PRCA with the BECN scheme. Although requiring a more dedicated rate control mechanism, the PRCA offers a more robust protection against loss of control cells, link failure, misbehaving users and abrupt traffic variations. Our study in the basic BECN scheme can be used as a guideline for specifying the system parameters in further investigation. In addition, the results suggest that a slow-start procedure, which allows traffic sources to reach a balanced cell rate in a conservative way (with the selection of appropriate control parameters, a similar effect can be achieved by using the PRCA), can be employed to reduce buffer requirements and cell loss.

Error control. The problem of cell discarding due to buffer overflow, excessive delay or bit errors is inherent in ATM networks. Cell discarding will severely degrade the service quality of real-time traffic since packet retransmission methods are not practical for this traffic type. One promising approach that is capable of reducing the quality degradation is to use cell loss recovery mechanisms such as Forward Error Correction (FEC) techniques. FEC involves transmission of redundant cells or frames in order to recover from losses without retransmissions. As real-time traffic such as video services becomes one of the significant factors driving the evolution of ATM networking technology, the demand for high-quality error control schemes is also increased. For the study of ATM error controls, the switch simulator and network models that we developed can serve as experimental platforms for simulating a realistic network environment.

REFERENCES

[1] Marek Wernik, Osama Aboul Magd and Henry Gilbert, "Traffic Management for B-ISDN Services," *IEEE Network,* vol.6, no.5, pp.10-19, Sept., 1992.

[2] William R. Byrne, George Clapp, Henry J. Kafka, Gottfried W.R. Luderer and Bruce L. Nelson, "Evolution of Metropolitan Area Networks to Broadband ISDN," *IEEE Communication Magazine,* vol.29, no.1 pp.69-82, Jan., 1991.

[3] M.R. Wernik and E.A. Munter, "Broadband Public Network and Switch Architecture," *IEEE Communication Magazine,* vol.29, no.1 pp.83-89, Jan., 1991.

[4] Edoardo Biagioni, Eric Cooper and Robert Sansom, "Designing a Practical ATM LAN," *IEEE Network,* vol.7, no.2, pp.32-39, March, 1993.

[5] Peter Newman, "ATM Technology for Corporate Networks," *IEEE Communication Magazine,* vol.30, no.4, pp.10-19, April, 1992.

[6] Gopalakrishnan Ramamurthy and Rajiv S. Dighe, "A Multidimensional Framework for Congestion Control in B-ISDN," *IEEE J. Select. Areas Commun.,* vol.9, no.9, pp.1440-1451, Dec., 1991.

[7] Zhixing Ren and James S. Meditch, " A Two-Layer Congestion Control Protocol for Broadband ISDN," in *Proc. INFOCOM '92,* April 1992, pp.0925-0933.

[8] Alexander Gersht and Kyoo J. Lee, "A Congestion Control Framework for ATM Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1119-1130, Sept., 1991.

[9] Bharat T. Doshi and Subrahmanyam Dravida, "Congestion Control for Bursty Data in High Speed Wide Area Packet Networks: In-Call Parameter Negotiations," *ITC Specialist Seminar 7,* Morristown, NJ, 1990.

[10] S. Jamaloddin Golestani, "Congestion-Free Communication in High-Speed Packet Networks," *IEEE Trans. Commun.,* vol.39, no.12, pp.1802-1812, Dec., 1991.

[11] Thomas M. Chen and Steve S. Liu, "Management and Control Functions in ATM Switching Systems," *IEEE Network,* vol.8, no.4, pp.27-40, July/August, 1994.

[12] M. Decina and T. Toniatti, "On Bandwidth Allocation to Bursty Virtual Connections in ATM Networks," in *Proc. ICC'90,* 1990, paper 318.6, pp.844-851.

[13] M. Decina, T. Toniatti, P. Vaccari, and L. Verri, "Bandwidth Assignment and Virtual Call Blocking in ATM Networks," in *Proc. INFOCOM '90,* April 1990, pp.881-888.

[14] G. Gallassi, G. Rigolio, and L. Fratta, "ATM: Bandwidth Assignment and Bandwidth Enforcement Policies," in *Proc. GLOBECOM '89,* 1989, paper 49.6, pp.1788-1793.

[15] G. M. Woodruff and R. Kositpaiboon, "Multimedia Traffic Management Principles for Guaranteed ATM Network Performance," *IEEE J. Select. Areas Commun.,* vol.8, no.3, pp.437-446, Apr., 1990.

[16] Roch Guerin, Hamid Ahmadi and Mahmoud Naghshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.968-981, Sept., 1991.

[17] Anwar I. Elwalid and Debasis Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE/ACM Trans. on Networking,* vol.1, no.3, pp.329-343, June, 1993.

[18] Bharat T. Doshi and Harry Heffes, "Performance of an In-Call Buffer-Window Reservation/Allocation Scheme for Long File Transfers," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1013-1023, Sept., 1991.

[19] Jonathan S. Turner, "New Directions in Communications (or Which Way to the Information Age?)," *IEEE Communication Magazine,* vol.24, no.10, pp.8-15, Oct., 1986.

[20] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks," in *Proc. INFOCOM '92,* April 1992, pp.915-924.

[21] Jonathan S. Turner, "Managing Bandwidth in ATM Networks with Bursty Traffic," *IEEE Network,* vol.6, no.5, pp.50-58, Sept., 1992.

[22] Gopalakrishnan Ramamurthy and Rajiv S. Dighe, "Distributed Source Control: A Network Access Control for Integrated Broadband Packet Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.990-1002, Sept., 1991.

[23] Lixia Zhang, "VirtualClock: A New Traffic Control Algorithm for Packet-Switched Networks," *ACM Trans. on Computer Systems,* vol.9, no.2, pp.101-124, May, 1991.

[24] S. Jamaloddin Golestani, "A Framing Strategy for Congestion Management," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1064-1077, Sept., 1991.

[25] F.A. Tobagi, "Fast Packet Switch Architecture for B-ISDN," *Proceedings of the IEEE,* vol.78, no.1, pp.133-167, Jan., 1990.

[26] Michael G. Hluchyj and Mark J. Karol, "Queueing in High-Performance Packet Switching," *IEEE J. Select. Areas Commun.,* vol.6, no.9, pp.1587-1597, Dec., 1988.

[27] Jeane S.-C. Chen, Roch Guerin, and Tomas E. Stern, "Markov-Modulated Flow Model for the Output Queues of a Packet Switch," *IEEE Trans. Commun.,* vol.40, no.6, pp.1098-1110, June, 1992.

[28] Ilias Iliadis, "Performance of a Packet Switch with Input and Output Queueing under Unbalanced Traffic," in *Proc. INFOCOM '92,* April 1992, pp.0743-0752.

[29] Farouk Kamoun and Leonard Kleinrock, "Analysis of Shared Finite Storage in a Computer Network Node Environment under General Traffic Conditions," *IEEE Trans. Commun.*, vol.28, no.7, pp.992-1003, July, 1980.

[30] D. Tipper and M. K. Sundareshan, "Adaptive Policies for Optimal Buffer Management in Dynamic Load Environments," in *Proc. INFOCOM '88,* April 1988, pp.0535-0544.

[31] Jaime Jungok Bae and Tatsuya Suda, "Survey of Traffic Control Schemes and Protocols in ATM Networks," *Proceedings of the IEEE,* vol.79, no.2, pp.170-189, Feb., 1991.

[32] Xian Cheng and Ian F. Akyildiz, "A Finite Buffer Two Class Queue with Different Scheduling and Push-Out Schemes," in *Proc. INFOCOM '92,* April 1992, pp.0231-0241.

[33] Partha P. Bhattacharya and Anthony Ephremides, "Optimal Scheduling with Strict Deadlines," *IEEE Trans. Automat. Contr.,* vol.34, no.7, pp.721-728, July, 1989.

[34] Renu Chipalkatti, James F. Kurose and Don Towsley, "Scheduling Policies for Real-Time and Non-Real-Time Traffic in a Statistical Multiplexer," in *Proc. INFOCOM '89,* April, 1989, pp.774-783.

[35] Hans Kroner, Gerard Hebuterne, Pierre Boyer and Annie Gravey, "Priority Management in ATM Switching Nodes," *IEEE J. Select. Areas Commun.,* vol.9, no.3, pp.418-427, April, 1991.

[36] Changhwan Oh, Masayuki Murata and Hideo Miyahara, "Priority Control ATM for Switching Systems," *IEICE Trans. Commun.,* vol.E75-B, no.9, pp.894-905, Sept., 1992.

[37] Youngho Lim and John Kobza, "Analysis of a Delay-Dependent Priority Discipline in a Multi-Class Traffic Packet Switching Node," in *Proc. INFOCOM '88,* April, 1988, pp.0889-0898.

[38] N. Yin, S-Q. Li, and T. E. Stern, "Congestion Control for Packet Voice by Selective Packet Discarding," in *Proc. IEEE GLOBECOM '87,* pp.45.3.1-45.3.4.

[39] David W. Petr, Luiz A. Dasilva, Jr., and Victor S. Frost, "Priority Discarding of Speed in Integrated Packet Networks," *IEEE J. Select. Areas Commun.,* vol.7, no.5, pp.644-656, June, 1989.

[40] S. Dravida and K. Sriram, "End-to-End Performance Models for Variable Bit Rate Voice over Tandem Links in Packet Networks," in *Proc. INFOCOM '89,* April 1989, pp.1089-1097.

[41] D. J. Goodman, "Embedded DPCM for variable bit rate transmission," *IEEE Trans. Commun.,* vol.28, pp.1040-1046, July, 1980.

[42] M. Devault, J. Cochennec, and M. Servel, "The Prelude ATD Experiment: Assessments and Future Prospects," *IEEE J. Select. Areas Commun.,* vol.6, no. 9, pp.1528-1537, Dec., 1988.

[43] H. Kuwahara, N. Endo, M. Ogino, and T. Kozaki, "Shared Buffer Memory Switch for an ATM Exchange," in *Proc. Int. Conf. on Communications*, Boston, MA, June 1989, pp.4.4.1-4.4.5.

[44] I. Cidon, *et al.*, "Real-time Packet Switching: a Performance Analysis," *IEEE J. Select. Areas Commun.*, vol.6, no.9, pp.1576-1586, Dec., 1988.

[45] P. Barri and J. A. O. Goubert, "Implementation of a 16 to 16 Switching Element for ATM Exchanges," *IEEE J. Select. Areas Commun.*, vol.9, no.5, pp.751-757, June, 1991.

[46] W. Fischer, O. Fundneider, E.-H. Goeldner, and K.A. Lutz, "A Scalable ATM Switching System Architecture," *IEEE J. Select. Areas Commun.*, vol.9, no. 8, pp.1299-1307, Oct., 1991.

[47] A. Itoh, W. Takahashi, H. Nagano, M. Kurisaka, and S. Iwasaki, "Practical Implementation and Packaging Technologies for a Large Scale ATM Switching System," *IEEE J. Select. Areas Commun.*, vol.9, no. 8, pp.1280-1288, Oct., 1991.

[48] I. Gard and J. Rooth, "An ATM Switching Implementation-Technique and Technology," *Proc. International Switching Symp. (ISS '90)*, vol.4, pp.23-27, 1990

[49] L.R. Goke and G.J. Lipovski, "Banyan Networks for Partitioning Multiprocessor Systems," *Proc. First Annual Symp. Comput. Architect*, pp.21-28, Dec., 1973.

[50] G.J. Anido and A.W. Seeto, "Multipath Interconnection: A Technique for Reducing Congestion Within Fast Packet Switching Fabrics," *IEEE J. Select. Areas Commun.*, vol.6, no. 9, pp.1480-1488, Dec., 1988.

[51] E.E. White, "A Quantitative Comparison of Architectures for ATM Switching System," *Washington University at Saint Louis*, WUCS-91-47,Sept., 1992.

[52] M.J. Karol, M.G. Hluchyj, and S.P. Morgan, "Input versus Output Queueing on a Space-division Packet Switch, " *IEEE Trans. Commun.*, vol.35, no.12, pp.1347-1356, Dec., 1987.

[53] Demers A., Keshav S., and Shenker S., "Analysis and Simulation of a Fair Queueing Algorithm," in *Proc. SIGCOMM'89*, Sept. 1989.

[54] Aramaki, Toshiya, *et al.*, "Parallel ATOM Switch Architecture for High-Speed ATM Networks," *IEEE International Conference on Communications, Supercomm/ICC* , pp.250-254 June, 1992.

[55] A. E. Eckberg, "B-ISDN/ATM Traffic and Congestion Control," *IEEE Network*, vol.6, no.5, pp.28-37, Sept., 1992.

[56] Wolfgang Fischer, Eugen Wallmeier, Thomas Worster, Simon P. Davis, and Andrew Hayter, "Data Communications Using ATM: Architectures, protocols, and Resource Management," *IEEE Communication Magazine*, vol.32, no.8, pp.24-32, Aug., 1994.

[57] Peter Newman, "Traffic Management for ATM Local Area Networks," *IEEE Commun. Magazine*, vol.32, no.8, pp.44-50, Aug., 1994.

[58] Brett J. Vickers and Tatsuya Suda, "Traffic Modeling for Telecommunications Networks," *IEEE Communication Magazine*, vol.32, no.3, pp.34-42, March, 1994.

[59] Victor S. Frost and Benjamin Melamed, "Connectionless Service for Public ATM Networks," *IEEE Communication Magazine*, vol.32, no.8, pp.70-81, Aug., 1994.

[60] T. Kamitake and T. Suda, "Evaluation of an Admission Control Scheme for an ATM Network considering fluctuations in cell loss rate," in *Proc. IEEE GLOBECOM '89,* pp.49.4.1-49.4.7.

[61] D. Heyman, A. Tabatabai, and T. V. lakshman, "Statistical Analysis and Simulation Study of Video teletraffic in ATM Networks," *IEEE Trans. Circuits and Systems for Video Technology,* vol.2, pp.49-59, 1992.

[62] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Trans. Commun.,* vol.36, pp.834-844, July, 1988.

[63] Y. Yasuda, H. Yasuda, N. Ohta, and F. Kishino, "Packet Video Transmission through ATM Networks," in *Proc. IEEE GLOBECOM '89,* pp.25.1.1-25.1.5.

[64] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic Theory of a Data-Handling System with Multiple Sources," *The Bell System Technical Journal,* vol.61, no.8, pp.1871-1894, 1982.

[65] D. L. Jagerman and B. Melamed, "The Transition and Autocorrelation Structure of TES Processes Part I: General Theory," *Stochastic Models,* vol.8, no.2, pp.193-219, 1992.

[66] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences,* John Wiley and Sons, New York, NY, 1979.

[67] W. E. Leland et al., "On the Self-Similar Nature of Ethernet Traffic," in *Proc. SIGCOMM '93,* pp.183-193, 1993.

[68] H.T. Kung, Robert Morris, Thomas Charuhas and Dong Lin, "Use of Link-by-Link Flow Control in Maximizing ATM Networks Performance: Simulation Results," in *Proc. IEEE Hot Interconnects Symposium'93,* Aug., 1993.

[69] David W. Petr and Victor S. Frost, "Optimized Nested Threshold Cell Discarding for ATM Overload Control," *International Journal of Digital and Analog Communication Systems,* vol.5, pp.97-116, 1992.

[70] Arthur Y.-M. Lin and John A. Silvester, "Priority Queueing Strategies and Buffer Allocation Protocols for Traffic Control at an ATM Integrated Broadband Switching System," *IEEE J.Select. Areas Commun.,* vol.9, no.9, pp.1524-1536, Dec., 1991.

[71] William H. Press, Brian P. Flannery, Saul A. Teukolsky and William T. Vetterling, *Numerical Recipes in C,* Cambridge University Press, London, UK, 1988.

[72] Jaime Jungok Bae, Tatsuya Suda and Rahul Simha, "Analysis of Individual Packet Loss in a Finite Buffer Queue with Heterogeneous Markov Modulated Arrival Processes: A Study of Traffic Burstiness and Priority Packet Discarding," in *Proc. INFOCOM '92,* April, 1992, pp.0219-0230.

[73] Hans Kroner, "Comparative Performance Study of Space Priority Mechanisms for ATM Networks," in *Proc. INFOCOM '90,* June, 1990, pp.1136-1143.

[74] J.F. Meyer, S. Montagna and R. Paglino, "Dimensioning of an ATM Switch with Shared Buffer and Threshold Priority," *Computer Networks and ISDN Systems,* 26, pp.95-108, 1993.

[75] B. T. Doshi and H. Heffes, "Overload Performance of Several Processor Queueing Disciplines for the M/M/1 Queue, " *IEEE Transactions on Communication,* vol.34, no.6, pp.538-546, June, 1986.

[76] G. Hebuterne and A. Gravey, "A Space Priority Queueing Mechaniam for Multiplexing ATM Channels," in *Proc. ITC Spec. Sem. '89,* Adelaide, Australia, Sept. 1989, paper 7.4.

[77] Allyn Romanow, "TCP over ATM: Some Performance Results," ATM Forum Contribution Report 93-784, July, 1993.

[78] Van Jacobson, "Congestion Avoidance and Control," in *Proc. SIGCOMM'88,* pp. 314-329, Aug., 1988.

[79] Lixia Zhang, Scott Shenker and David D. Clark, "Observations on the Dynamics of a Congestion Control Algorithm: The Effects of Two-Way Traffic," in *Proc. SIGCOMM'91,* pp. 133-146, Aug., 1991.

[80] Andrea Bianco, "Performance of the TCP Protocol over ATM Networks," in *Proc. ICCCN'94,* pp. 170-177, Sept., 1994.

[81] H.T. Kung, Trevor Blackwell and Alan Chapman, "Credit-Based Flow Control for ATM Networks: Credit Update Protocol, Adaptive Credit Allocation, and Statistical Multiplexing," in *Proc. SIGCOMM'94,* Aug., 1994.

[82] Peter Newman, "Simulation Results for a Backward Explicit Congestion Control Scheme," ANSI T1S1.5/93-047, Raleigh NC, Feb., 1993.

[83] Peter Newman, "Backward Explicit Congestion Notification for ATM Local Area Networks," in *Proc. GLOBECOM'93,* pp. 719-723, Nov., 1993.

[84] V. Jacobson, R. Braden, D. Borman, "TCP Extensions for High Performance," *Request for Comments (RFC) 1323*, May 1992.

[85] Andrew W. Barnhart, "Baseline Performance Using PRCA Rate-Control," ATM Forum Contribution Report 94-0597, July 1994.

[86] Hiroshi Saito and Kohei Shiomoto, "Dynamic Call Admission Control in ATM Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.982-989, Sept., 1991.

[87] Mark J. Karol and Salman Z. Shaikh, "A Simple Adaptive Routing Scheme for Congestion Control in ShuffleNet Multihop Lightwave Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1040-1051, Sept., 1991.

[88] Ellen L. Hahne, "Round-Robin Scheduling for Max-Min Fairness in Data Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1024-1039, Sept., 1991.

[89] Jay M. Hyman, Aurel A. Lazar and Giovanni Pacifici, "Real-Time Scheduling with Quality of Service Constraints," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1052-1063, Sept., 1991.

[90] B. Ngo and H. Lee, "Queueing Analysis of Traffic Access Control Strategies with Preemptive and Nonpreemptive Disciplines in Wideband Integrated Networks," *IEEE J. Select. Areas Commun.,* vol.9, no.7, pp.1093-1109, Sept., 1991.

[91] Nanying Yin and Michael G. Hluchyj, "A Dynamic Rate Control Mechanism for Source Coded Traffic in a Fast Packet Network," *IEEE J. Select. Areas Commun.*, vol.9, no.7, pp.1003-1012, Sept., 1991.

[92] Jeane S.-C. Chen and Tomas E. Stern, "Throughput Analysis, Optimal Buffer Allocation, and Traffic Imbalance Study of a Generic Nonblocking Packet Switch," *IEEE J. Select. Areas Commun.*, vol.9, no.3, pp.439-449, April, 1991.

[93] Milena Butto, Elisa Cavallero and Alberto Tonietti, "Effectiveness of the Leaky Bucket Policing Mechanism in ATM Networks," *IEEE J. Select. Areas Commun.*, vol.9, no.3, pp.335-342, April, 1991.

[94] Yoshihiro Ohba, Masayuki and Hideo Miyahara, "Analysis of Interdeparture Processes for Bursty Traffic in ATM Networks," *IEEE J. Select. Areas Commun.*, vol.9, no.3, pp.468-476, April, 1991.

[95] Geert A. Awater and Frits C. Schoute, "Optimal Queueing Policies for Fast Packet Switching of Mixed Traffic," *IEEE J. Select. Areas Commun.*, vol.9, no.3, pp.458-467, April, 1991.

[96] Domenico Ferrari and Dinesh C. Verma, "A Scheme for Real-Time Channel Establishment in Wide-Area Networks," *IEEE J. Select. Areas Commun.*, vol.8, no.3, pp.368-379, April, 1990.

[97] Kim-Joan Chen, Kelvin K.Y. Ho and Vikram R. Saksena, "Analysis and Design of a Highly Reliable Transport Architecture for ISDN Frame-Relay Networks," *IEEE J. Select. Areas Commun.*, vol.7, no.8, pp.1231-1242, Oct., 1989.

[98] F. Kishino, K. Manabe, Y. Hayashi and H. Yasuda, "Variable Bit-Rate Coding of Video Signals for ATM Networks," *IEEE J. Select. Areas Commun.*, vol.7, no.5, pp.801-806, June, 1989.

[99] Thomas M. Chen, Jean Walrand and David G. Messerschmitt, "Dynamic Priority Protocols for Packet Voice," *IEEE J. Select. Areas Commun.*, vol.7, no.5, pp.632-643, June, 1989.

[100] Joseph Y. Hui, "Resource Allocation for Broadband Networks," *IEEE J. Select. Areas Commun.*, vol.6, no.9, pp.1598-1608, Dec., 1988.

[101] Israel Cidon, Inder Gopal, George Grover and Moshe Sidi, "Real-Time Packet Switching: A Performance Analysis," *IEEE J. Select. Areas Commun.*, vol.6, no.9, pp.1576-1586, Dec., 1988.

[102] Harry Heffes and David M. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE J. Select. Areas Commun.*, vol.4, no.6, pp.856-868, Sept., 1986.

[103] Raj Jain and Shawn A. Routhier, "Packet Train - Measurements and a New Model for Computer Network Traffic," *IEEE J. Select. Areas Commun.*, vol.4, no.6, pp.986-995, Sept., 1986.

[104] Kotikalapudi Sriram and Ward Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE J. Select. Areas Commun.*, vol.4, no.6, pp.833-846, Sept., 1986.

[105] Janusz Filipiak, "Accuracy of Traffic Modeling in Fast Packet Switching," *IEEE Trans. Commun.*, vol.40, no.4, pp.835-846, April, 1992.

[106] Debasis Mitra, "Asymptotically Optimal Design of Congestion Control for High Speed Data Networks," *IEEE Trans. Commun.*, vol.40, no.2, pp.301-311, Feb., 1992.

[107] San-Qi Li, "Performance of a Nonblocking Space-Division Packet Switch with Correlated Input Traffic," *IEEE Trans. Commun.*, vol.40, no.1, pp.97-108, Jan., 1992.

[108] Nedo Celandroni and Erina Ferro, "The FODA-TDMA Satellite Access Scheme: Presentation, Study of the System, and Results," *IEEE Trans. Commun.*, vol.39, no.12, pp.1823-1831, Dec., 1991.

[109] Jaidev Kaniyil, Yoshikuni Onozato, Ken Katayama and Shoichi Noguchi, "Input Buffer Limiting: Behavior Analysis of a Node Throughout the Range of Blocking Probabilities," *IEEE Trans. Commun.*, vol.39, no.12, pp.1813-1822, Dec., 1991.

[110] Kerry W. Fendick, Vikram R. Saksena and Ward Whitt, "Investigating Dependence in Packet Queues with the Index of Dispersion for Work," *IEEE Trans. Commun.*, vol.39, no.8, pp.1231-1244, Aug., 1991.

[111] San-Qi Li, "A General Solution Technique for Discrete Queueing Analysis of Multimedia Traffic on ATM," *IEEE Trans. Commun.*, vol.39, no.7, pp.1115-1132, July, 1991.

[112] Samuel P. Morgan, "Queueing Disciplines and Passive Congestion Control in Byte-Stream Networks," *IEEE Trans. Commun.*, vol.39, no.7, pp.1097-1106, July, 1991.

[113] Ken W. Sarkies, "The Bypass Queue in Fast Packet Switching," *IEEE Trans. Commun.*, vol.39, no.5, pp.766-774, May, 1991.

[114] Riccardo Melen and Jonathan S. Turner, "Distributed Protocols for Access Arbitration in Tree-Structured Communication Channels," *IEEE Trans. Commun.*, vol.39, no.3, pp.416-425, March, 1991.

[115] Jeane S.-C. Chen and Roch Guerin, "Performance Study of an Input Queueing Packet Switch with Two Priority Classes," *IEEE Trans. Commun.*, vol.39, no.1, pp.117-126, Jan., 1991.

[116] C. Murray Woodside and Eric D. S. Ho, "Engineering Calculation of Overflow Probabilities in Buffers with Markov-Interrupted Service," *IEEE Trans. Commun.*, vol.35, no.12, pp.1272-1277, Dec., 1987.

[117] Mark J. Karol, Michael G. Hluchyj and Samuel P. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Trans. Commun.*, vol.35, no.12, pp.1347-1356, Dec., 1987.

[118] John B. Nagle, "On Packet Switches with Infinite Storage," *IEEE Trans. Commun.*, vol.35, no.4, pp.435-438, April, 1987.

[119] Tien-Yu Huang, Jean-Lien Chen Wu and Jingshown Wu, "Priority Management to Improve the QOS in ATM Networks," *IEICE Trans. Commun.*, vol.E76-B, no.3, pp.249-257, Mar., 1993.

[120] Zvi Rosberg and Parviz Kermani, "Customer Scheduling under Queueing Constraints," *IEEE Trans. Automat. Contr.,* vol.37, no.2, pp.252-257, Feb., 1992.

[121] Steve H. Lu and P. R. Kumar, "Distributed Scheduling Based on Due Dates and Buffer Priorities," *IEEE Trans. Automat. Contr.,* vol.36, no.12, pp.1406-1416, Dec., 1991.

[122] Costas A. Courcoubetis and Martin I. Reiman, "Optimal Control of a Queueing System with Simultaneous Service Requirements," *IEEE Trans. Automat. Contr.,* vol.32, no.8, pp.717-727, Aug., 1987.

[123] Fouad A. Tobagi, Mario Gerla, Richard W. Peebles and Eric G. Manning, "Modeling and Measurement Technigues in Packet Commmunication Networks," *Proceedings of the IEEE,* vol.66, no.11, pp.1423-1447, Nov., 1978.

[124] Rene L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Trans. Information Theory,* vol.37, no.1, pp.114-131, Jan., 1991.

[125] Rene L. Cruz, "A Calculus for Network Delay, Part II: Network Analysis," *IEEE Trans. Information Theory,* vol.37, no.1, pp.132-141, Jan., 1991.

[126] James W. roberts, "Variable-Bit-rate traffic Control in BISDN," *IEEE Commun. Magazine,* vol., no., pp.50-56, Sept., 1991.

[127] Ljiljana Trajkovic and S. Jamaloddin Golestani, "Congestion Control for Multimedia Services," *IEEE Network,* vol.6, no.5, pp.20-26, Sept., 1992.

[128] H. Heffes, "A Class of Data Traffic Processes-Covariance Function Characterization and Related Queueing Results," *The Bell System Technical Journal,* July-August, pp.897-929, 1980.

[129] Young Han Kim, Byung Chul Shin and Chong Kwan Un, "Performance Analysis of Leaky-bucket Bandwidch Enforcement Strategy for Bursty Traffics in an ATM Network," *Computer Networks and ISDN Systems,* 25, pp.295-303, 1992.

[130] Juha Heinanen, "Frame Relay as a Multiprotocol Backbone Interface," *Computer Networks and ISDN Systems,* 25, pp.363-369, 1992.

[131] Otto Koudelka and Mervyn Hine, "Higher Speed Services," *Computer Networks and ISDN Systems,* 16, pp.129-134, 1988/89.

[132] Jim Kurose, "Open Issues and Challenges in Providing Quality of Service Guarantees in High-Speed Networks," *Computer Communication Review,* ACM SIGCOMM, vol.25, no.1, pp.6-15, Jan. 1993.

[133] Opher Yaron and Moshe Sidi, "Calculating Performance Bounds in Communication Networks," in *Proc. INFOCOM '93,* April 1993, pp.539-546.

[134] L.K. Reiss and L.F. Merakos, "Shaping of Virtual Path Traffic for ATM B-ISDN," in *Proc. INFOCOM '93,* April 1993, pp.168-175.

[135] Leandros Tassiulas, Yaochung Hung and Shivendra S. Panwar, "Optimal Buffer Control during Congestion in an ATM Network Node," in *Proc. INFOCOM '93,* April 1993, pp.1059-1066.

[136] G. Ramamurthy and B. Sengupta, "A Predictive Hop-by-Hop Congestion Control Policy for High Speed Networks," in *Proc. INFOCOM '93,* April 1993, pp.1033-1041.

[137] F. Bernabei, L. Gratta, M. Listanti and A. Sarghini, "Analysis of On-Off Source Shaping for ATM Multiplexing," in *Proc. INFOCOM '93,* April 1993, pp.1330-1336.

[138] Ji Zhang, "Performance Study of Markov Modulated Fluid Flow Models with Priority Traffic," in *Proc. INFOCOM '93,* April 1993, pp.10-17.

[139] Hiroshi Toyoizumi, "A Simple Method of Estimating Mean Delay by Counting Arrivals and Departures," in *Proc. INFOCOM '93,* April 1993, pp.829-834.

[140] Bart Steyaert, Herwig Bruneel and Yijun Xiong, "A General Relationship between Buffer Occupancy and Delay in Discrete-Time Multiserver Queueing Models, Applicable in ATM Networks," in *Proc. INFOCOM '93,* April 1993, pp.1250-1258.

[141] Chung G. Kang and Harry H. Tan, "Queueing Analysis of Explicit Priority Assignment Partial Buffer Sharing Schemes for ATM Networks," in *Proc. INFOCOM '93,* April 1993, pp.810-819.

[142] San-Qi Li and Song Chong, "Fundamental Limits of Input Rate Control in High Speed Network," in *Proc. INFOCOM '93,* April 1993, pp.662-671.

[143] D. Tipper, J. Hammond, S.Sharma, A. Khetan, K. Balakrishnan and S. Menon, "An Analysis of the Congestion Effects of Link Failures in Wide Area Networks," in *Proc. INFOCOM '93,* April 1993, pp.1042-1050.

[144] Israel Cidon and Roch Gueerin, "On Protective Buffer Policies," in *Proc. INFOCOM '93,* April 1993, pp.1051-1058.

[145] Soung C. Liew and Tony T. Lee, "A Fundamental Property for Traffic Management in ATM Networks," in *Proc. INFOCOM '93,* April 1993, pp.1240-1249.

[146] Hyeog-In Kwon, Arnon Tubtiang and Guy Pujolle, "A Simple Flow Control Mechanism in ATM Network with End to End Transport," in *Proc. INFOCOM '93,* April 1993, pp.654-661.

[147] Christoph Herrmann, "Analysis of the Discrete-time SMP/D/1/s Finite Buffer Queue with Applications in ATM," in *Proc. INFOCOM '93,* Apri l 1993, pp.160-167.

[148] Roch Guerin and Levent Gun, "A Unified Approach to Bandwidth Allocation and Access Control in Fast Packet-Switched Networks," in *Proc. INFOCOM '92,* April 1992, pp.0001-0012.

[149] David X. Chen and Jon W. Mark, "A Buffer Management Scheme for the SCOQ Switch under Nonuniform Traffic Loading," in *Proc. INFOCOM '92,* April 1992, pp.0132-0140.

BIOGRAPHICAL SKETCH

Shang-Yi Lu was born on January 1, 1964, in Taiwan, R.O.C. She graduated from Kao-Shiung High School in 1982. She then went to Shin-Chu, a city 50 miles from Taipei, entered the National Chiao-Tung University, and graduated with a Bachelor of Science degree in electrical engineering in 1986. She joined China Technical Consultants Inc. in 1986 as an electrical engineer responsible for planning and evaluation of communication and switching systems. She came to the USA in 1988 for advanced study in the electrical engineering program at the University of Missouri-Columbia, resulting in a Master of Science degree in 1990. She then transferred to the University of Florida for the electrical engineering Ph.D. program, where she has been a research assistant and has been involved with several research projects in the Laboratory for Information Systems and Telecommunications.

She was a scholarship winner from the National Chiao-Tung University. She is currently a student member of the Institute of Electrical and Electronics Engineers (IEEE).

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Haniph A. Latchman , Chairman
Associate Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Leon W. Couch II
Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald G. Childers
Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Scott Miller
Associate Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Randy Chow
Professor of Computer and Information Sciences

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1994

Winfred M. Phillips
Dean, College of Engineering

Karen A. Holbrook
Dean, Graduate School